



EDINBURGH  
UNIVERSITY  
LIBRARY

Shelf Mark Theses Section 2

RENTON, Ph.D. 2000



AN APOLOGY FOR MATERIALISM

ALASTAIR RENTON

THESIS SUBMITTED FOR THE DEGREE OF PH.D  
UNIVERSITY OF EDINBURGH

1999



DECLARATION OF AUTHORSHIP

I DECLARE THIS TO BE MY OWN WORK;

*Alastair Kenton*

## ABSTRACT

It is natural to suppose that mental and physical properties are importantly distinct. Yet whatever this difference is, it has to be compatible with interaction between the mind and the body. Satisfaction of these desiderata leads to a paradox. If you make the mind strongly separate from the body, then there is the problem of bringing them together. If you unite them, then there is the problem of preserving their distinctiveness. It is the aim of this thesis to resolve the paradox.

From the outset, it is assumed that the nature of interaction is most satisfactorily explained by an account of mental properties in monistic terms. For reasons of space, the arguments of Materialism are concentrated upon at the expense of Idealism. Three strategies are examined, and found wanting. First, an instance of a non-reductive account provided by Davidson's 'Anomalous Monism'. Here, mental properties seem to be left with no rôle in influencing behaviour. Second, a review of reductionist accounts, ranging from Identity Theories to Representationalism. Criticism focuses upon the failure of reductionism to explain the connection between the function of a conscious state and its particular character.

A Materialist treats mental states as if they were part of the physical universe. This implies that the nature of these states may be understood through scientific investigation, in the same manner as all other phenomena. The third strategy is to deny the above implication: that is, deny the assertion that, by existing, all aspects of an object are thereby knowable. The ideas of Colin McGinn are discussed as an example of this position. Since his arguments are equally suitable for non-Materialist purposes, they do not constitute an exclusively Materialist solution to the above paradox.

This thesis offers an alternative way of pursuing the above strategy. It argues that the relation between mental states and our ways of understanding phenomena, is such that we should not expect our theories about the nature of 'mind' and the 'physical world' to employ the same terms. These properties appear distinct, not because they are different substances, but because they occupy different sides of the 'process of understanding' - 'thing understood' relationship. For convenience, this position is referred to as 'Agnostic Materialism'.

As interaction between the mind and the body is compatible with the mind having no influence upon our behaviour, it is incumbent upon the thesis to defend Materialism against the claim that mental properties are epiphenomenal. This is achieved by teasing out two ways in which such properties are considered inert: either because the workings of the mind are independent of the body; or because the mind's processes are irrelevant to those of the body. The first claim is seen arise from the difficulty of seeing the mind as part of the physical world - a difficulty removed by the arguments in the previous paragraph. The second claim gains plausibility through a mistaken adherence to certain models of scientific explanation.



## Table of Contents

<u>Chapter One</u>	1
Section One	1
Subsection 1 : Why try to resolve the paradox?	1
Section Two	2
Subsection 1 : How is this resolution to be achieved?	2
Subsection 2 : Materialist, Physicalist or Naturalist?	2
Section Three	3
Subsection 1 : The character of sensations.	4
Subsection 2 : The intentionality of thoughts.	5
Subsection 3 : The sense of personal identity.	6
Subsection 4 : Conclusion.	7
Section Four	7
Subsection 1 : The problem of Epiphenomenalism.	7
Subsection 2 : The problem of Determinism.	8
Section Five	9
Section Six	10
Subsection 1 : Caveat.	10
Subsection 2 : Plan of campaign.	10
<u>Chapter Two</u>	13
Section One	15
Subsection 1 : Bringing the three principles into line.	16
Subsection 2 : Psycho-physical connections and strict laws.	17
Subsection 3 : Davidson's Identity theory.	18
Section Two	20
Subsection 1 : Does Anomalous Monism entail 'Type Epiphenomenalism'?	21
Subsection 2 : Extensionalism as a defence against Epiphenomenalism.	24
Subsection 3 : Supervenience as a defence against Epiphenomenalism.	26
Subsection 4 : Non-strict laws as a defence against Epiphenomenalism.	29
Subsection 5 : Has enough been said about the character of sensations?	31
Section Three	31
Subsection 1 : General problems with Supervenience.	31
Subsection 2 : Horgan's remarks upon Supervenience.	33
Subsection 3 : Wilson's amendment to 'Superdupervenience'.	35
Subsection 4 : Conclusion.	37

Section Four	38
Subsection 1 : Crane - some explanations are better than others.	38
Subsection 2 : Sosa - the drawbacks to treating all causes equally.	39
Subsection 3 : Conclusion.	40
 <u>Chapter Three</u>	 42
Section One	43
Subsection 1 : Behaviourism.	43
Subsection 2 : Identity theories.	44
 Section Two	 46
Subsection 1 : Are absent qualia possible?	47
Subsection 2 : Are inverted qualia possible?	48
Subsection 3 : Implications of the above possibilities for Functionalism.	48
 Section Three	 49
Subsection 1 : Sober - Functions as 'purposes'.	49
Subsection 2 : Lycan - 'Homuncular Functionalism'.	50
Subsection 3 : 'Function' as a Teleological concept.	52
Subsection 4 : Functionalist solutions to the qualia problems.	54
 Section Four	 55
Subsection 1 : Can Functionalism deal with the 'inverted qualia'?	56
Subsection 2 : Does Functionalism explain why we have sensations?	57
 Section Five	 58
Subsection 1 : The appeal of Representationalism.	59
Subsection 2 : The relation between Functionalism and Representationalism.	60
 Section Six	 61
Subsection 1 : Is philosophy the correct discipline for understanding the mind?	62
Subsection 2 : Are there reasonable grounds for fearing Eliminativism?	63
Subsection 3 : Can we live without Naturalising the mind?	64
Subsection 4 : Response to Stich.	65
 Section Seven	 65
Subsection 1 : The notion of two levels of consciousness.	67
Subsection 2 : Rosenthal's HOT theory of consciousness.	68
Subsection 3 : Tye's HOT theory of consciousness.	71
Subsection 4 : Lycan's HOE theory of consciousness.	73
Subsection 5 : General problems with vertical theories of consciousness.	74
Subsection 6 : Conclusion.	75
 <u>Chapter Four</u>	 76
Section One	76
Subsection 1 : The Horizontal theory of consciousness.	76

Subsection 2 : Problems with the 'Internalist Intuition'.	77
Subsection 3 : A defence of the 'Internalist Intuition'.	79
Subsection 4 : Consciousness as a product of Evolution.	80
Subsection 5 : Problems with this account.	82
Subsection 6 : The Externalist element in Horizontal theories.	83
Subsection 7 : Problems with this analysis of consciousness.	85
 Section Two	 87
Subsection 1 : The origin of the conflict between Externalism and Internalism.	87
Subsection 2 : Egan - there is no useful distinction between these positions.	89
Subsection 3 : Walsh - resolving the conflict by appeal to context.	92
Subsection 4 : Kirk - problems with Ultra-Externalism.	93
Subsection 5 : Is Externalism consistent with Representationalism?	94
 Section Three	 97
Subsection 1 : How intelligible is the possibility of inverted qualia?	98
Subsection 2 : Block - the intrinsic nature of qualia.	101
Subsection 3 : Has the riddle of inversion been solved?	104
 Section Four	 105
 <u>Chapter Five</u>	 109
 Section One	 110
Subsection 1 : Nagel – could there be an objective account of 'subjectivity'?	110
Subsection 2 : Jackson and 'the knowledge argument'.	113
 Section Two	 115
Subsection 1 : Levine's remarks on conceivability arguments.	116
Subsection 2 : Levine and the 'explanatory gap'.	120
 Section Three	 122
Subsection 1 : Why the psycho-physical link is a natural property of the brain.	123
Subsection 2 : Cognitive closure and the CALM hypothesis.	123
Subsection 3 : Attempts to discover the nature of the link via introspection.	125
Subsection 4 : A 'reductio' to prove that the link is essentially hidden.	126
Subsection 5 : Attempts to discover the nature of the link via perception.	127
Subsection 6 : Learning to live with 'Cognitive Closure'.	129
Subsection 7 : Conclusion.	129
 Section Four	 129
Subsection 1 : Is McGinn's Naturalism compatible with his Agnosticism?	129
Subsection 2 : Does McGinn make unreasonable demands of science?	130
Subsection 3 : Problems with McGinn's 'reductio' argument.	131
Subsection 4 : Does McGinn's position redeem Materialism?	133

<u>Chapter Six</u>	135
Section One	136
Subsection 1 : A brief introduction to the terms ‘cause’ and ‘explanation’.	136
Subsection 2 : Levine on the rôle of properties in causal explanations.	138
Subsection 3 : Implications for a Materialist analysis of mental causation.	139
Section Two	140
Subsection 1 : Kripke’s remarks upon identifying mental with physical states.	141
Subsection 2 : Implications of the above for the viability of Science.	142
Subsection 3 : Implications for the viability of Dualism.	143
Subsection 4 : Is Science compatible with a contingent mind-body relation?	145
Subsection 5 : Conclusion.	146
Section Three	147
Subsection 1 : The rôle of sensations in scientific progress.	147
Subsection 2 : The puzzle of understanding the nature of sensations.	148
Section Four	150
Subsection 1 : An analysis of theory formation.	150
Subsection 2 : A brief sketch concerning the nature of the emotions.	152
Subsection 3 : A difficulty for Materialist theories of emotion and a solution.	154
Subsection 4 : Nagel and problems with a Materialist analysis of the ‘self’.	155
Subsection 5 : Conclusion.	158
Section Five	158
Subsection 1 : Is Agnosticism amenable to a Dualist interpretation?	158
Subsection 2 : Is Agnosticism amenable to an Idealist interpretation?	160
Subsection 3 : Conclusion.	161
<u>Chapter Seven</u>	163
Section One	165
Subsection 1 : Is consciousness overrated?	166
Subsection 2 : Scientific support for Weak <sub>E</sub> .	167
Subsection 3 : The Zombie possibility.	169
Section Two	170
Subsection 1 : Defence of Agency on the basis of explanatory importance.	171
Subsection 2 : Response - inconsistency in our veneration of patterns.	172
Subsection 3 : Is Epiphenomenalism an unprofitable form of scepticism?	174
Section Three	175
Subsection 1 : Kirk and the logical possibility of Zombies.	175
Subsection 2 : Problems with Kirk’s account.	176
Subsection 3 : How not to imagine a Zombie - as a counterfeit.	177

Subsection 4 : Are Zombies nomologically possible?	178
Subsection 5 : The impossibility of imagining myself to be a Zombie.	178
Subsection 6 : An alternative reason for accepting Strong <sub>E</sub> .	180
Subsection 7 : The inconsistency of these intuitions.	181
Subsection 8 : The problem of analysing 'cause'.	182
Subsection 9 : Is the Principle of autonomy at fault?	184
Subsection 10 : Conclusion.	184
 Section Four	 185
Subsection 1 : Is consciousness irrelevant?	186
Subsection 2 : Steward's criticism of the 'Network' model of causation.	187
Subsection 3 : Cummins and a critique of models of explanation.	188
Subsection 4 : Problems with trusting scientific models of explanation.	189
Subsection 5 : Is consciousness redundant?	190
Subsection 6 : The importance of 'feelings' to our choices of action.	191
Subsection 7 : The importance of 'perceptions' to our choices of action.	192
Subsection 8 : Is it a contingent truth that the mind is not Epiphenomenal?	193
 Section Five	 193
Subsection 1 : Searle and emergent properties.	193
Subsection 2 : Lowe's criticism, and re-interpretation of emergent properties.	194
Subsection 3 : Difficulties with Lowe's position.	195
Subsection 4 : Alternative ways of thinking about causes.	196
Subsection 5 : Conclusion.	198
 <u>Chapter Eight</u>	 199
 Section One	 199
Subsection 1 : Summary.	199
Subsection 2 : Disclaimer.	199
 Section Two	 200
Subsection 1 : Implications for topics in moral philosophy.	200
Subsection 2 : Implications for topics in epistemology.	201
 <u>Bibliography</u>	 203

## Chapter One

How insecure do you have to be, before you start wondering whether the mind has any influence on the body? Many people would reply: "Very". This connection is blithely taken for granted. There is, as yet, no justification for this nonchalant attitude. But, each attempt to provide such a justification, has met with - at best - disagreement, or - at worst - sniggers of derision. Therefore, a wiser response to the above question would be: "Hardly at all."

Two ideas appear superficially uncontroversial: the way that I am informed about the state of my mind is importantly different from the way in which I am informed about the rest of the world: states of mind affect and are affected by states of the world. The difficulties begin once one tries to unite these two beliefs.

It is natural to suppose that the nature of an object of enquiry, determines the method of enquiry. If it is inappropriate to use the same method to discover facts about both the mind and the world, what does this say about the natures of these entities? It suggests that the mind and the world are two distinct substances.

One might also assume that if two things are able to influence each other, then they must have some common properties. Where things have properties in common, they cannot be truly considered as being composed of distinct substances.

Therefore, our ideas about the mind and the world result in a paradox. If we hold to the distinctness of the mind, then it is difficult to see how it interacts with the world. If we hold to the interaction between the mind and the world, then it is difficult to account for how our knowledge of our mind and the world is so different. It seems that we cannot remain constant to one idea without thereby being unfaithful to the other. This thesis is one more attempt to resolve the above paradox.

### Section One

#### Subsection 1 : Why try to resolve the paradox?

It is possible to live with the paradox, and regard it as being a quirk of our intellect, rather than as an indicator of something truly life-threatening. Yet, one reason why a person would be bothered by it, is because of the problem it raises about the relation between the mind and the world: does the mind play any part in the natural course of events? If we believe that through science we will eventually gain a complete understanding of the nature of the world, then there will have to be a place for the mind in the great scientific catalogue of 'what there is'. Unfortunately, whether or not the mind is composed of the same substance as everything else, it is not obvious how some mental states are to be described physically. Therefore it is difficult to see how the mind and its workings could be accommodated within the scientific scheme of things. This is especially ironic; that the seat of understanding - the mind - should

be denied a place in its own creation. In some sense, this is a clue to resolving the paradox.

## Section Two

### Subsection 1 : How is this resolution to be achieved?

There are at least two ways of resolving the paradox. One is to affirm the existence of two substances, and thereby guarantee the intuitive separateness of the mind. This leaves the problem of interaction; but there is no logical sense in which it is impossible for two distinct substances to affect each other, however mysterious the means might be. The other is to affirm the existence of only one substance, and thereby hope that the tale of interaction will be easier to tell. This leaves the problem of explaining the discreteness of the mind. Somehow, it has to be shown that the problems with identifying mental states with physical states, does not count against this strategy. Neither path is conspicuously straightforward, but I have chosen a version of the latter, since my grasp of science is marginally in better shape than my appreciation of the nature of the immaterial.

### Subsection 2 : Materialist, Physicalist or Naturalist?

The first problem for the Materialist, is to sort out whether or not they are also Physicalists or Naturalists. Though these terms are by no means equivalent, they share a commitment to ontological simplicity. A Materialist believes that everything that exists is fundamentally composed of material particles. Physicalists accept this principle, and hold that it is the rôle of Physics to discover and classify the properties of these particles, and the laws that constrain their behaviour. There are a number of important criticisms of this position, that suggest it is too restrictive an account<sup>1</sup>. Naturalism may be regarded as a weaker version of Physicalism, holding that the Natural Sciences generally, should be employed to do this task. For the purposes of this thesis, it is assumed that Materialism commits one to the existence of a single type of substance, that has properties which determine the ways in which parts combine and react; that the multifarious scientific disciplines take this substance as its subject matter, whether it be physics, geology or psychology. The only sense of reduction that is adhered to, is the sense that all phenomena are composed from this material. For reasons that will emerge, whilst believing in the existence of the totality of facts, there is a commitment to our knowledge of these facts being a proper subset of such a totality. These terms are only used interchangeably where the context is one where the difference between the positions is irrelevant. Usually, this reflects an author's usage of the term; its presence in the text is not intended to indicate an overall endorsement of that particular position.

---

<sup>1</sup> See Crane and Mellor 1990, Sturgeon 1998, for examples of these criticisms.

### Section Three

How can we avoid denying that mind is different from matter? The mind does not readily conform to our expectations of material things, and our concepts reflect this. The entry for 'Mind' in the New Shorter Oxford English Dictionary, includes this definition:

The seat of awareness, thought, volition, and feeling; cognitive and emotional phenomena and powers as constituting a controlling system, specifically as opposed *matter*; the spiritual as distinguished from the bodily part of a human being.<sup>2</sup>

The following is a quote from David Lewis, exemplifying a persuasive account of the nature of the world:

This world, or any possible world, consists of things which instantiate fundamental properties and which, in pairs or triples or..., instantiate fundamental relations...It is a task of Physics to provide an inventory of all the fundamental properties and relations that occur in the world...We have no a priori guarantee of it, but we may reasonably think that present-day physics goes a long way toward a complete and correct inventory.<sup>3</sup>

It is clear from the definition of 'Mind', that it is traditionally contrasted with matter. Perhaps the most appealing consideration for thinking that the mind as a whole is not to be understood in physical terms, appears in the 'Discourse on the Method', by Rene Descartes:

...I saw that while I could feign that I had no body, that there was no world, and no place existed for me to be in, I could not feign that I was not; on the contrary, from the mere fact that I thought of doubting about other truths it evidently and certainly followed that I existed. On the other hand, if I had merely ceased to be conscious, even if everything else that I had ever imagined had been true, I had no reason to believe that I should have existed. From this I recognised that I was a substance whose whole essence or nature is to be conscious and whose being requires no place and depends upon no material thing. Thus this self, that is to say the soul, by which I am what I am, is entirely distinct from the body, and is even more easily known; and even if the body were not there at all, the soul would be just what it is.<sup>4</sup>

Whatever the philosophic merits of this passage, it is faithful to common belief: the mind is an entity that eludes physical classification, in virtue of its apparent independence from anything physical. In particular, there seem to be three distinct regions where the 'traditional' understanding of our minds bulges over the sides of the scientific framework, like a thirty-six inch waist in a pair of size twenty-eight jeans. To continue the metaphor, the choice is between regarding these aspects of the mind as so much unsightly

---

<sup>2</sup> New Shorter Oxford English Dictionary. 1993. P 1779.

<sup>3</sup> Lewis. 1994. P 412.

<sup>4</sup> Descartes. 1954. P 32.



and embarrassing fat, or feeling quite happy about it, and buying larger - ontological - trousers.

The following Subsections discuss the three features of the mind that resist our attempts to colonise them in the name of science.

### Subsection 1 : The character of sensations.

Whilst 'Consciousness' is not a direct synonym of 'Mind', it is clearly very much a mental concept. Recently, through the work of Timothy Sprigge in "Final Causes", and Thomas Nagel in "What is it like to be a Bat?", consciousness has been characterised as follows:

One is wondering about the consciousness which an object possesses whenever one wonders what it must be like being that object.<sup>5</sup>

In our assessment of whether or not an entity is conscious, one could resort to this definition. The problem is, how to know when one has successfully appreciated what - if anything - it is like to be that object?

When one imagines another's conscious state, there is no conclusive way of checking up whether one has done so correctly or not. This by no means implies that one's guess may not in fact be more or less correct. Presuming that the object...with which one is concerned, is indeed conscious, then *being that organism* will have a certain definite complex quality at every walking moment...Physical science makes no reference to qualities of this kind.<sup>6</sup>

These "complex qualities" are familiar enough in our own case, be they thoughts or states of sensory experience. Yet, this familiarity is not so intimate as to banish the following doubts: how are such qualities produced; are they *really* present in other creatures? Nor does it appear likely that scientific understanding, however complete, will bring about such an intimacy. Thus Frank Jackson challenges the Materialist:

Tell me everything physical there is to tell about what is going on in a living brain, the kinds of states, their functional role, their relation to what goes on at other times and in other brains, and so on and so forth, and be I as clever as can be in fitting it all together, you won't have told me about the hurtfulness of pains, the itchiness of itches, pangs of jealousy, or about the characteristic experience of tasting a lemon, smelling a rose, hearing a loud noise or seeing the sky.<sup>7</sup>

Two related issues arise from this. Suppose there is a complete explanation for our behaviour that leaves it undetermined which mental states are associated with those actions. Is it possible

---

<sup>5</sup> Sprigge. 1971. P 167.

<sup>6</sup> Sprigge. 1971. P 168.

<sup>7</sup> Jackson. 1982. P 127.

that the same physical description could be true of two people, that they behave in the same way, but have different types of experience? Is it possible that there could be creatures who lack consciousness altogether, whilst acting in a 'normal' manner? If the mind is physical, then there should be straightforward answers to these questions. A moment's reflection is enough to show that knowing where to look for these answers is going to be a hard task.

### Subsection 2 : The intentionality of thoughts.

Geoffrey Madell makes the following point about intentionality:

If we are to pay attention to the phenomenology of experience, these two factors, the non-dispositionality and the irreducibility of the directedness of thought, seem absolutely undeniable.<sup>8</sup>

Most of our thoughts, are directed towards something; that is, they have intentional content. This idea of 'Intentionality' is notoriously hard to capture in words, but it is a property that we could not coherently deny exists. Any denial worthy of the name, would consist of thoughts with intentional content. Doubtless a person could think complete gibberish, and thereby have thoughts that were completely lacking in intentional content; certainly, there are some species of imaginings which lack any particular object. Whether or not these observations make it any easier to explain thoughts in physical terms, is beside the point. For those thoughts that *are* intentional, there is the problem of finding the physical property that corresponds to this feature.

Some philosophers have suggested that 'thought' would be more amenable to physical description if it were treated as a type of disposition. Any species of thought, be it a belief, or a desire, for instance, can be understood in terms of how the person entertaining that thought is disposed to act. No doubt this is roughly accurate of how we attribute thoughts to others, but there is an obvious difficulty when we apply this method in the attribution of thoughts to ourselves. Madell observes:

...if my thinking of something is a matter of my being disposed to behave in some way, it must follow that I can have no idea what it is that I am thinking of until this behavioural disposition is manifested.<sup>9</sup>

Consider the following instance of an act of thinking. I have just been reading a passage from Siegfried Sassoon's "Memoirs of an Infantry Officer". I pause, and I - as it were - say to myself: "My Grandfather was based in the same part of France; I wonder whether Sassoon knew him?" On the dispositional account, the thought is constituted by its disposing me to act in certain ways. However, to have any knowledge of my behaviour, it is necessary to form beliefs

---

<sup>8</sup> Madell. 1988. P 11.

<sup>9</sup> Madell. 1988. P 12.

concerning the state of the world. These beliefs themselves, being thoughts, would have to be construed as dispositions, and so on; the implausible conclusion of this analysis of thought would have me dying without raising the above query concerning my Grandfather.

Since it has become possible to instantiate logical processes in machines, it is tempting to conclude that there must be a physical analysis of 'intelligence'<sup>10</sup>. Yet, 'intelligence' is not at issue, but states of mind. Such an analysis would provide an understanding of how it is possible to behave intelligently, but not necessarily how we, as persons, manage to behave intelligently. Whether we think like machines, is a question yet to be answered. It is arguable that these physical processes are regarded as 'intentional' only in a derivative sense, by humans judging the performance of the machine. It is an essential feature of human intentionality that one is aware of it immediately rather than derivatively.

Intentionality fails to fit into the pattern of physical explanation. The only reason we have for not shrugging our shoulders and saying; "So much for Intentionality", is that without it, there is no making sense of; "...our immediate, non-inferential knowledge of our thoughts and occurrent emotions."<sup>11</sup> The Materialist must find some way of capturing this phenomenon.

### Subsection 3 : The sense of personal identity.

We have an irreducible sense of identity, no matter how difficult it is to articulate (for it cannot be denied that we lack, as yet, an uncontroversial account of the nature of personal identity). This awareness we have of ourselves as an individual, has to be accommodated within a Materialist framework. Thomas Nagel argues that any attempt to do so is bound to end in failure:

Given a complete description of the world from no particular point of view, including all the people in it, one of whom is Thomas Nagel, it seems on the one hand that something has been left out, something absolutely essential remains to be specified, namely which of them I am. But on the other hand there seems no room in the centreless world for such a further fact.<sup>12</sup>

We have no difficulty in individuating other people from a physical description, because in some sense, that is the only way we have of doing it. Nor do we have any difficulty individuating ourselves physically, such as when we see photographs or films of ourselves. The difficulty arises as soon as we accept that there is a physical description available for every process. For, if this is so, then there is such a description that corresponds with the way I normally individuate myself. Nagel's point is that as soon as you attempt to objectify such knowledge, then it ceases to be the phenomenon of self-awareness. Self-knowledge has an essentially

---

<sup>10</sup> See Alan Turing's article for an influential discussion of this possibility. Turing. 1950.

<sup>11</sup> Madell. 1988. P 131.

<sup>12</sup> Nagel. 1986. Pp 54-55.

perspectival character - a subjectivity that is incompatible with the objectivity of scientific knowledge.

#### Subsection 4 : Conclusion.

The foregoing considerations will have to be accommodated by anyone tempted by the possibility of bringing all aspects of our experience within one ontological kind. Otherwise, one is led to conclude that the best explanation for what there is, lies in there being two distinct and independent substances; that which underlies mental states, and is known through the activity of introspection, and that which underlies physical states and is investigated by the natural sciences.

### Section Four

We have a fair knowledge of how Nature works. We also have a fair knowledge of how our minds work. What we lack, is an uncontroversial understanding of how the two combine. Until that moment of enlightenment, we are prone to a pair of dark suspicions: that the mind has no influence at all: that the mind has influence, but no choice. No Materialist analysis of the mind can be considered complete without addressing the problems of Epiphenomenalism and Determinism.

#### Subsection 1 : The problem of Epiphenomenalism.

The following is a common conception of the rôle of science in explaining human action, provided by Norman Malcolm:

Neurological states and processes are conceived to be correlated by general laws with the mechanisms that produce movements. Chemical and electrical changes in the nervous tissue of the body are assumed to cause muscle contractions, which in turn cause movements such as blinking, breathing, and puckering of the lips, as well as movements of fingers, limbs and head.<sup>13</sup>

It is striking, as Malcolm notes, that these laws make: "...no provision for desires, aims, goals, purposes, motives or intentions."<sup>14</sup> One of the features of purposive explanation is that it cannot be assimilated to explanation in physical terms. This is a consequence of those aspects of the mind that were discussed above in Section Three. Malcolm enumerates three ways in which physical and purposive explanations differ. It is suggested firstly, that the former belong to a systematic and comprehensive set of statements comprising a theory, and that these elements are absent in purposive explanations. Secondly, that the concepts of 'purpose' and 'intention' do not appear in physical theories. Thirdly, that the former employ contingent laws which purposive explanations do not. If purposive explanation cannot be combined with physical

---

<sup>13</sup> Malcolm. 1982. P 127.

<sup>14</sup> Malcolm. 1982. P 128.

explanations, then they cannot be employed in our overall picture of what there is. However, if we are confident that there are physical explanations covering all there is, then this immiscibility is of no consequence to our knowledge:

On this view our ordinary explanations of behaviour will often be true: but the neural explanations will also be true - and they will be *more fundamental*. Thus we could, theoretically, *by-pass* explanations of behaviour in terms of purpose, and the day might come when they simply fall into disuse.<sup>15</sup>

There is a danger that if we believe that there are explanations for everything in physical terms, then there is no room for the purposive explanations to do any work. This leads to the feeling that referring to psychological explanations will become redundant; that the mental properties will not be making a difference

### Subsection 2 : The problem of Determinism.

Determinism is a view that becomes increasingly harder to resist as scientific discovery advances. Our understanding of nature is based on there being regularities between causes and effects. A more sophisticated account of this relation between events, recognises that, either there is complete randomness between two events, or the preceding event makes the succeeding event more likely to occur. In either case, it is plausible that the events that constitute our behaviour conform to this description. This presents us with a dilemma. If we treat behaviour as part of the natural order, then we must doubt the existence of free will. If we place our will outside the limits of the natural order, then we must doubt the possibility of being able to provide a physical description of our unconstrained choices. This line of thought appears in Peter van Inwagen's article 'The Incompatibility of Free Will and Determinism'. Determinism is described as being the conjunction of two theses:

- (a) For every instant of time, there is a proposition that expresses the state of the world at that instant.
- (b) If *A* and *B* are any propositions that express the state of the world at some instants, then the conjunction of *A* with the laws of physics entails *B*.<sup>16</sup>

Van Inwagen defines a law of physics, as a law of nature that is independent of the voluntary behaviour of rational agents. The notion of free will is captured thus:

To deny that men have free will is to assert that what a man *does* do and what he *can* do coincide. And almost all philosophers agree that a necessary

---

<sup>15</sup> Malcolm. 1982. P 131.

<sup>16</sup> van Inwagen. 1982. P 47.

condition for holding an agent responsible for an act is believing that that agent *could have* refrained from performing that act.<sup>17</sup>

The incompatibility of these two seemingly innocuous suppositions, arises from the following consideration. Suppose proposition 'P<sub>1</sub>' expresses a truth about the state of the world in the past. In conjunction with the laws of physics, it entails the proposition 'P<sub>2</sub>', expressing a truth about the state of the world now. Suppose that my current action 'A', is part of the proposition 'P<sub>2</sub>'. If 'A' is to be described as 'free', then I could have refrained from doing it: that is, I could have performed a different act 'B'. Van Inwagen points out, that the performance of 'B' instead of 'A' would falsify the truth of proposition 'P<sub>2</sub>'. This contradicts the belief that 'P<sub>2</sub>' is entailed by 'P<sub>1</sub>' together with the laws of physics. Either we must abandon our faith in physics, or our belief in free will; neither option strikes us as at all tenable.

### Section Five

If only we could refer to a mental state using the vocabulary of science, then we would be able to appreciate how the mind meshed with the non-mental workings of the brain. This does not seem to be an extravagant wish. Science has successfully replaced folklore as a means of providing an understanding of phenomena. In the past, objects of mystery have become analysed as the predictable outcomes of patterns of physical behaviour. Thus, flashes of lightning have been identified as electrical discharges, instead of weapons of Divine vengeance. Why should it not be the case that the coloured humours of our soul become identified with some physical state of our bodies? We now think emotions that would once have been attributed to 'black bile', may now be associated with chemical imbalances. What is to prevent Science from discovering the identity between the humour and the hormone?

The simple reason lies with the nature of the identity relation. Kripke argued that any identity between referring terms must hold of necessity. This has the following consequence for attempts to identify brain states with mental states:

Let 'A' name a particular pain sensation, and let 'B' name the corresponding brain state, or the brain state some identity theorist wishes to identify with A. *Prima facie*, it would seem that it is at least logically possible that B should have existed...without Jones feeling any pain at all, and thus without the presence of A.<sup>18</sup>

Consider an attempt to identify the mental state 'pain' with C-fibre activity in the brain. If the identity holds, then, necessarily, pain is the firing of C-fibres. However, this conclusion is one that seems wholly implausible. Note, it is not being claimed that the

---

<sup>17</sup> van Inwagen. 1982. Pp 49-50.

<sup>18</sup> Kripke. 1972. P 146.

activity would *cause* pain. That in itself would be counterintuitive enough; but it would point to a relation of cause and effect, rather than one of identity. Instead, the implication is that whenever there is a pain, it is a firing of C-fibres; and whenever there is a firing of C-fibres, it is painful. This conclusion contradicts the natural thought that, not only is it the case that pain could be experienced even though this tissue were inactive or absent, but equally, C-fibre firing could occur at a time when the individual felt nothing.

If this is what we think is true, then it is inappropriate to identify any mental state type with any brain state type:

Someone who wishes to maintain an identity thesis cannot simply *accept* the Cartesian intuitions that *A* can exist without *B*, that *B* can exist without *A*, that the correlative presence of anything with mental properties is merely contingent to *B*, and that the correlative presence of any specific physical properties is merely contingent to *A*.<sup>19</sup>

Any materialist will have to show that these intuitions are just an illusion, which as Kripke rightly notes, is not child's play.

## Section Six

### Subsection 1 : Caveat.

There is insufficient space to do more than acknowledge the fact that there are other ways of providing a Monistic account of the Mind: Idealism has been a particularly influential alternative to Materialism in this respect. Whilst there are many varieties of Idealism, they share a commitment to a division between the mind and reality, whereby reality is in some way dependent upon and fashioned by mental states and processes. Without judging whether any form of Idealism could be tenable, it is supposed that the remarks in this thesis make the mind part of reality. Certainly, it is accepted here that what one learns about reality, has to be in terms compatible with the processes of human understanding. However, this is only to acknowledge an epistemological relation between mind and reality: it is assumed here that reality is metaphysically independent of the mind.

### Subsection 2 : Plan of campaign.

A Materialist might reasonably assume that it is possible to provide a wholly satisfactory account of all phenomena employing some sort of natural vocabulary – leaving it open for now whether or not this will be accomplished by science. When it comes to understanding mental phenomena, Materialists typically adopt one of two approaches. Either they will deny that mental states are reducible to physical states, though there is a supervenience relation between the two which justifies their being identified. This position has the promise of conferring a distinctive quality upon mental

---

<sup>19</sup> Kripke. 1972. P 148.

states – in accordance with our intuition that the mind is in some way separate from the body. Or they will affirm that mental states are nothing more than physical states – that the properties of mental states can be seen as deriving from the properties of matter. This position respects the other intuition, that in explaining the relation between mind and body, we need look no further than everyday causal relations between physical properties.

Is there a need for yet another contribution to the debate on the nature of consciousness? In order to justify the effort required in providing an alternative to the many positions available, the first part of this thesis is negative in character. The next three chapters examine attempts to provide both non-reductive and reductive analyses of the mind. Chapter Two is a survey of Anomalous Monism where the link between the mind and the body is explained in terms of a Supervenience relation. Chapters Three and Four are concerned with the development of Representationalism. Both examples have attracted severe criticism. Both have trouble giving a convincing account of the relevance of mental properties. This is because they do not provide a physically satisfying analysis of mental properties – especially conscious states – in physical terms. Until these analyses are forthcoming, it seems inevitable that Materialism will be susceptible to the epistemological attacks made by ‘knowledge’ and ‘conceivability’ arguments.

There are many attractive suggestions to be found in the above positions, concerning how to construe cognitive processes in physical terms. Yet, the principle by which the physical and the mental might be connected, remains an enigma. Joseph Levine has christened the gulf between our knowledge of matter and of consciousness, the ‘Explanatory Gap’. Until this gap is bridged, or has been shown to be compatible with a Materialist point of view, Materialism cannot be considered to be a tenable position.

An alternative approach, is to question the assumption offered at the beginning of this Section; that all phenomena can be accounted for in terms of a natural vocabulary. Instead, one may affirm the following. The fact that we are aware of the existence of an object, provides no reason for thinking that we should be in a position to provide a complete, natural, description of that object. This position admits that the gap is unbridgeable, but that this is a result of there being categories of facts that are beyond our comprehension: in particular, that category of facts concerning the means by which the mind and the body could be composed of the same substance. A version of this line of thought will be defended below.

One way of articulating this position, is to claim that there are naturally occurring lacunae in our understanding: it is, after all, a plausible expectation that our intellect has its limits. Chapter Five considers McGinn’s work in this area. However, one chief criticism is that such a strategy is equally compatible with a Substance Dualism. This being so, it is not distinctively Materialist.



What emerges from these chapters, is that the Materialist has a two-fold problem; to close the explanatory gap, and to account for how the three categories of mind - sensation, intentionality and selfhood - can be reconciled with a Materialistic vocabulary. The sixth Chapter argues that these categories are an essential part of any information processing system. As such, they stand in a relation to our understanding, as to make them appear to be of a different category from other objects of experience. To this extent, it is compatible with Representationalist theories. However, it seeks to overcome the problem of the nature of sensation, by arguing that the nature of the physical basis of sensation - and mind-brain interaction generally - is ineffable, as a consequence of the relation between the processes of understanding and the objects of understanding. The processes themselves cannot be understood in quite the same way as the objects, and these differences are borne out in the notable dissimilarities between our understanding of our minds as opposed to the world. Thus the problem of identifying mental and physical states, arises as a result of the distinctive ways in which we are acquainted with them, rather than because they are of separate substances. Whilst they are separate from an epistemic point of view, from a metaphysical perspective, the mind and the body belong to the same category.

Chapter Seven uses this conclusion to show how the problem of interaction can be avoided by Materialists. To overcome the objection that science leaves no room for mental properties, it is argued that such a conclusion depends upon certain models of explanation. These are demonstrably unreliable; a putative explanation may conform to one of these models, and yet misrepresent the causal relations between the properties cited in the explanans. There is a tendency to believe that an explanation has been provided once you have picked-out a property P connecting adjacent events. This assumes that P is responsible for such a sequence, and thereby overlooks the following possibility. P might be inert, and the causal efficacy lies in a further property - unknown to us - with which P is co-instantiated. Therein lies the space where mental states, in virtue of their physical properties, to which we have no epistemic access, can make a difference to the world.

Finally in Chapter Eight, the thesis concludes with brief remarks concerning the impact of these conjectures upon other philosophical issues.

## Chapter Two

If we are to be justified in holding that our beliefs are arrived at by a process that is not determined by natural laws, then we must accept that there are distinct laws concerning mental processes and natural processes. That is to say, if our belief in free will is to be more than just an illusion, then the mind must be able to make a difference to what occurs in the world. Herein lies the problem of mental causation. If our understanding of the goings-on of the world is to be complete, then we will find ourselves with two distinct sets of laws: one governing natural events, the other governing mental events. These have to be such as to make it intelligible that the mind affects, and is affected by, natural events, even though our intuitions suggest that the connections between sequences of mental states are intrinsically different from those between natural states. Paradoxically, the mind has to be autonomous with respect to nature, and yet be sufficiently compatible to have an effect upon the course of natural events. From the outset, the idea of mental causation forces upon us both the need to separate the mental and physical, as well as the need to bring them together.

This chapter examines a non-reductivist solution to this problem: the position developed by Donald Davidson that has become known as 'Anomalous Monism'<sup>1</sup>. In "Mental Events", he alludes to the observation made by Immanuel Kant:

Philosophy must therefore assume that no true contradiction will be found between freedom and natural necessity in the same human actions, for it cannot give up the idea of nature any more than that of freedom.<sup>2</sup>

The motivation for Davidson's initial enquiry can be seen as attempting to reconcile these two convictions. This position is far too complex to be summarised within one chapter, so for the purposes of this thesis, the intention is to examine what consequences follow from such an account of the mind-body relationship. Accordingly, the first Section of this chapter will give an outline of the position, as stated in the article "Mental Events" published in 1970, and drawing from other papers of that period. These will be supplemented with Davidson's response<sup>3</sup> to criticisms made over the years following the 1970 publication. No attempt during this Section will be made to assess Davidson's remarks: such an assessment will form subsequent Sections of this chapter.

An important consideration unites many of the criticisms; there is a danger that the mental becomes Epiphenomenal if it is articulated in a manner consistent with Anomalous Monism. Arguably, the assumptions that are required to make free will compatible with the existence of strict laws, sit awkwardly with the assumptions required to make mental events causally relevant to

---

<sup>1</sup> Though Davidson could not be considered a Physicalist, it would not be libellous to regard his Monism as being a form of Materialism.

<sup>2</sup> As cited in Davidson 1980. P 207. Kant. 1909. P 75.

<sup>3</sup> Davidson. 1993.

physical events, and vice versa. The problem facing anyone hopeful of defending Anomalous Monism, is that there has to be a way of saying how the mental makes a difference; and more importantly, makes a difference that justifies our talk of 'mental causation'. These criticisms will be addressed in Section Two. A related problem for this position is generated by the use of the concept of 'supervenience' to relate the mental with the physical. Just how illuminating and appropriate it is to employ this concept, is examined in Section Three. Finally, there is a difficulty that arises from Davidson's exclusion of 'properties' in favour of particular events, when these are employed in explanations: Section Four deals with the relationship between the existence of properties and the possibility of explanation, suggesting that Anomalous Monism has an attenuated explanatory capacity.

These are not intended as an exhaustive list of the possible objections to the position. Rather, they are selected to highlight one of the difficulties listed in the first chapter - the obscure nature of the causal relation between the mental and the physical. The conclusion to be drawn from these remarks is that, for the monist, this relation is doomed to be elusive whilst it is accepted that the mental is in some sense nomologically distinctive - something upon which Davidson's arguments depend.

Before discussing Davidson's views in detail, there are a number of points concerning the subject matter that require clarification: for as Davidson observes, much of the criticism of his position is a result of failing to appreciate the ambiguity of usage.

An 'event' is a concrete particular:

Events are taken to be unrepeatable, dated individuals such as the particular eruption of a volcano, the (first) birth or death of a person, the playing of the 1968 World series, or the historic utterance of the words, 'You may fire when ready, Gridley.'<sup>4</sup>

Any particular event may be described as a mental event, such as; 'Deciding to read Gorky's "Life of a Useless Man", provided that the description is true. The same event could also be described as physical, such as; 'A period of excitation in region psi-4 of the occipital lobe'. Thus, it is quite consistent with Davidson's ideas that the same event can be characterised both as a mental event and as a physical event.

However there is an important difference to how we characterise the relationship between successive events described physically, and when described mentally. Treated as physical events, the succession may be thought of as being one event causing the other, according to some strict law. Here a strict law is distinguished from a generalisation, by being without exceptions; whereas a generalisation merely marks out an interesting pattern. It is assumed that the nearest examples of strict laws that we have, are those belonging to Physics. However, the pattern of mental events is

---

<sup>4</sup> Davidson. 1980. Pp 209-210.

somewhat different, following the principles of rationality rather than contingency. The implication of this distinction is that, so far as Davidson is concerned, there are no strict laws governing the relationship between mental events, nor applying to relations between the mental and the physical. Certainly, there are general laws which may be applied, but these as such do not warrant the ascription of 'causes' to events. Events are 'causes' or 'effects' only as they are described according to the vocabulary of some science that utilises 'strict laws', the paradigm of which is Physics.

This is not to imply that mental events cannot be causes. The terms 'cause' and 'effect' should be understood as playing an explanatory rôle here. To say that mental events cannot be causes, because only physical descriptions can, is to ignore the following condition: when one event follows on upon another, it does so regardless of the ways in which those events may be described. If a causal explanation is called for, then the appropriate medium to provide such an explanation, is in the vocabulary of Physics. In this respect, it would be incorrect to enquire whether mental events can be causes, since there is no such thing as a mental event per se; rather there are events that can be referred to using a mental vocabulary. The expressions 'cause' and 'effect' are reserved for explanations employing the vocabulary of a completed science.

Finally, it should be remembered that the mental states from which Davidson draws his examples, are all instances of propositional attitudes, such as 'desiring' and 'believing'. He is not directly concerned with those states which are associated with the senses, such as sensations of light, sound, smell, taste or touch. It is his belief that any remarks he makes about the propositional attitudes will be applicable to the senses.

### Section One

When discussing the relation between mind and body, Davidson suggests that two independent factors become crucial in identifying the stance one is prepared to take. One factor concerns the existence of law-like connections between the mental and the physical - whether the relationship between the two is nomological or anomalous. The other factor concerns the identification of the mental with the physical - whether the position is monistic or dualistic. The characterisation of any particular philosophical position will depend upon which commitments to the truth of these factors such a position embraces. Accordingly, there are four possible stances available: to affirm both the existence of psycho-physical laws and identity; to deny both; to accept psycho-physical laws but deny identity; to deny psycho-physical laws but accept identity. Of these four, Davidson accepts the last. He denies the existence of psycho-physical laws - hence the 'Anomalous' - but holds that there is an identity between mental and physical - hence the 'Monism'.

The motivation behind 'Mental Events' is revealed in the following query:

Mental events such as perceivings, rememberings, decisions, and actions resist capture in the nomological net of physical theory. How can this fact be reconciled with the causal role of mental events in the physical world? <sup>5</sup>

Here are two intuitions that he assumes are true; that our choices must make a difference in the way of the world, and that the order of the world is made perspicuous by strict laws. The question raised above is that these intuitions appear to conflict with one another. To clarify where the tension lies, he describes three Principles:

a) The Principle of Causal Interaction: there are some instances of mental events that are capable of interacting causally with physical events.

b) The Principle of the Nomological Character of Causality: in order for events to be related as cause and effect, they must be brought under strict deterministic laws.

c) The Principle of the Anomalism of the Mental: there are no such laws that can be used to predict or explain mental events.

On the face of it, there appears to be an inconsistency between these principles. The remainder of 'Mental Events' is divided into three parts: firstly, to suggest ways in which these principles can be brought together; secondly to show that there cannot be causal connections between the mental and the physical which fall under Principle 'b'; and finally to propose a version of the Identity theory that supports the identification between at least some mental and physical events.

### Subsection 1 : Bringing the three principles into line.

In order to relieve the tension between these principles, Davidson needs to show how it is possible for the mental to make a difference, whilst avoiding the implications of determinism. The way he achieves this, is to identify the mental with the physical at the level of events.

These remarks appear to satisfy the second and third principles, but the principle of causal interaction as it stands is not straightforwardly upheld by the event identity strategy Davidson uses. There remains a problem of how events described as 'mental' can have an influence on events described as 'physical', when causal influence is reserved for descriptions at a physical level.

The solution to this quandary, is to introduce the notion of 'Supervenience' in order to explicate the way in which the mental affects the physical. Davidson notes that supervenience was used to explain how evaluative judgements, such as 'good', 'wicked' and 'beautiful' could be applied to natural objects, without there being a commitment to saying that evaluative properties could be reduced to natural properties. For instance, to state that an evaluative property

---

<sup>5</sup> Davidson. 1980. P 207.

like 'mediocrity' supervenes upon a physical state, such as a book, commits the speaker to affirming of any physically identical book, that it was mediocre too. Supervenience is a dependence relation, such that a change in one level entails a change in the other. It does not entail that like changes at a one level will bring about like changes at the other - merely that no change can take place in isolation. Typically, the levels are referred to as higher and lower levels, where the higher level 'supervenes' upon the lower, or 'subvenient base', level.

Davidson adapts this idea to account for the interdependence between the mind and the body, where the mental is the supervenient level, and the physical its subvenient base:

Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect.<sup>6</sup>

In this manner, he is able to account for identity and causality between the mental and the physical, without falling foul of determinism. The three principles are thereby reconciled; the problem has been diagnosed as taking 'event' to be unambiguous when used in the expressions 'mental event' and 'physical event'. On Davidson's analysis, events are sensitive to how they are described. Similarly, there is no straightforward use of the term 'relation' that holds between events: the relation between two events described physically, will be as cause and effect; between two events described mentally, they will be related holistically; and between an event described mentally and one described physically, they will be related by the former supervening upon the latter. Once these niceties have been cleared up, the tension between the principles becomes relieved.

### Subsection 2 : Psycho-physical connections and strict laws.

The second part of 'Mental Events' defends the idea that there are no strict psycho-physical laws. We are faced with the problem of finding this idea both obviously true and palpably false. On the one hand, the failure of the Behaviourists to reduce definitions of mental concepts to behavioural ones, suggests that such laws are a fiction. Whenever a behavioural description was offered, it always had an unanalysed psychological component<sup>7</sup>. Yet on the other hand, the idea denies the existence of relations between mind and body that we use every day.

There seems to be an essential difference between explanations that use mental terms from those that use physical ones. It is a commonplace remark that sciences are 'objective'; but for all that, the point underlines the distinction between the accounts we give for physical and mental phenomena. Rightly or

---

<sup>6</sup> Davidson. 1980. P 214.

<sup>7</sup> For further remarks on this topic, see Chapter 3 below, Pp 43-44.

wrongly, we assume that scientific explanation is neutral according to the psychological state of the person giving the explanation. That is to say, the truth of science is established by uncovering patterns of change between entities according to their properties. The existence of these patterns is unaffected by whether or not they are understood. This is not a feature of psychological explanation. Here, an understanding of any particular mental event, will be affected by a complex relation between that person's mental states and dispositions. It is this feature of the relationship between mental states that prevents such states falling under strict laws:

Beliefs and desires issue in behaviour only as modified and mediated by further beliefs and desires, attitudes and attendings, without limit. Clearly this holism of the mental realm is a clue both to the autonomy and to the anomalous character of the mental.<sup>8</sup>

Davidson suggests the following formulation for what he accepts law-like to mean. He holds that there is a parallel between nomological and analytic, insofar as both are linked to meaning. As such, there will be constraints upon the types of predicate that can be linked together in a law. According to these constraints, mental and physical predicates are not of a compatible nature, and as such, cannot figure together in a law. This is not to deny that they cannot feature in generalisations; but generalisations cannot be refined sufficiently to qualify as laws in Davidson's sense. Whenever an attempt is made to reduce a mental term to a physical one, it invariably has to employ terms from a mental vocabulary, contrary to Davidson's definition of a law.

Drawing these two observations together, Davidson concludes:

It is a feature of physical reality that physical change can be explained by laws that connect it with other changes and conditions physically described. It is a feature of the mental that the attribution of mental phenomena must be responsible to the background of reasons, beliefs and intentions of the individual. There cannot be tight connections between the realms if each is to retain allegiance to its proper source of evidence.<sup>9</sup>

### Subsection 3 : Davidson's Identity theory.

By the final part of the article, Davidson is in a position to draw on the results of the previous parts, to produce two novel conclusions. Firstly, contrary to prevailing opinion, it is possible to produce an identity theory that does not require the existence of psycho-physical laws - in fact, which denies their existence. Secondly, that the grounds for accepting such an identity theory, are based upon both the anomalous nature of the mental and the nomological nature of the physical. The argument for identifying mental events with physical events runs as follows.

---

<sup>8</sup> Davidson. 1980. P 217.

<sup>9</sup> Davidson. 1980. P 222.

Assume that there are at least some causal relations between mental and physical events, both where mental causes have physical effects, and where physical causes have mental effects. Assume also that the truth of any singular causal statement, is warranted by appeal to strict laws, which connect events of the kind referred to in that statement. Such laws exist in the physical sciences. It is the nature of our mental concepts that they cannot be brought together under such laws to characterise the relations between mental events; nor are there any psycho-physical laws. So it would appear that mental phenomena are ineligible for nomological classification.

From these assumptions, the identification of mental events with physical events follows thus. Consider a mental event 'M', and a physical event 'P'. Where we say, "'M' caused 'P'", there will be a strict law covering instances of some description of 'P' kinds of events following 'M' kinds of events. This will be the case only if 'M' has a physical description. If it has, then 'M' must be identical with a physical event. Importantly, an event may be referred to by any of the mental or physical properties true of it; in contrast, an event is a cause or an effect only according to how its physical properties relate to strict laws.

Finally Davidson returns to the original question - how can the mental events be causally efficacious, in spite of the fact that they cannot be ordered according to any nomological schema worthy of the name? The answer is that the attributes 'mental' and 'physical' apply only to descriptions of events, rather than to sorts of causes. Causation can only be understood to hold between events as described physically, since this is the appropriate form for the formulation of strict laws. So to say that mental events have causes that produce physical effects, is to say that there is a relation of cause and effect between two events, and that the prior event has a mental description, whereas the latter event has a physical description. Importantly, the justification for citing that event 'A' caused 'B', is conferred by their being related to each other according to some strict physical law. Whilst events are described as mental, they fall out of the jurisdiction of physical inevitability. Those same events, described as physical, allow for the propriety in saying that mental events make a difference in the world:

When we portray events as perceivings, rememberings, decisions and actions, we necessarily locate them amid physical happenings through the relation of cause and effect; but as long as we do not change the idiom that same mode of portrayal insulates mental events from the strict laws that can in principle be called upon to explain and predict physical phenomena.<sup>10</sup>

The attraction of this means of relating the mind to the body, is that it retains identity, but avoids reduction. One of the daunting obstacles to monism has been the obligation to account for mental concepts in physical terms. Insofar as Anomalous Monism is a theory about how events are related, it is possible to retain the autonomy of the mind at the level of description, without the

---

<sup>10</sup> Davidson. 1980. P 225.



problem of having to explain how every mental concept is reducible to a physical one. Where the identity lies, is at the level of events, not at the level of concepts.

The proper explanation of the causal relation of events is given in physical terms, but this does not entail that there are no mental causes, so far as Davidson is concerned. An event is neither mental nor physical except as so described. If 'A' causes 'B', then there will be an interpretation under some physical law as to why that sequence of events took place. However, it would be as correct to describe 'A' in mental terms as physical, and, depending upon the context of the explanation, more informative. For as Davidson notes, it is important to keep separate the idea of the practice of causal explanation, from the idea of causal relations; the former are interest sensitive, whereas the latter are not. The speaking in a squeaky voice will have an explanation at the level of physiology, and one at the level of psychology. Both explanations are true in some sense, but the truth of their respective claims is underwritten in different ways. That is why Davidson considers the mental to be anomalous: its laws are simply of a different kind from those that are utilised in physics. As such, there cannot be a relation between the two - just as there cannot be a merging between the personal rules a player might adopt whilst playing chess (like always moving a pawn after your opponent coughs), with the rules that govern the playing of that game.

## Section Two

The possibility that the mental properties of an event may be Epiphenomenal, emerges when one considers, more generally, the causal relations between states of affairs. For instance, whilst driving along in your car, you hear a repetitive tapping sound. It turns out that the cause of the noise is a stone, trapped between the treads of the tyre, hitting the road surface each time the wheel rotates. It is natural to suppose that there are some properties of that stone which do not contribute to the noise - such as its colour, or that it was originally from an area outside Dagenham - in contrast with other properties which are required for the noise - such as its durability and size relative to the tyre's tread. It is generally accepted that not every property of an object will be relevant to the causal relations in which that object is involved. One abiding problem in the Philosophy of Mind, is to determine whether or not mental properties are ever causally efficacious. That is, in terms of the above example, whether the mental properties are like the hardness of the stone, or whether they are like the colour. If they are like the latter, then mental properties are Epiphenomenal.

It is obvious from Davidson's first principle, that he assumes that there is some sort of commerce between the mental and the physical. It is certain that he denies that the mind is Epiphenomenal. Yet many critics claim that Anomalous Monism entails Epiphenomenalism. Their criticism turns upon the coherence of the three principles. They question whether it is possible to hold

simultaneously, that events are individuated as causally related according to some strict law, that there are no strict psychological laws, and yet mental events causally interact with physical events. On the face of it, the supporter of Anomalous Monism must show how their idea of a causal relation can be compatible with the non-strict law-like nature of the mind.

This Section deals with possible lines of defence available to the Anomalous Monist. The first Subsection deals with the claim that Anomalous Monism entails one version of Epiphenomenalism, because of its adherence to the strict law thesis. This claim is arguably too strong. No-one yet has come up with a convincing argument to show that, as it happens, the property of falling under a strict law is the only property that confers causal efficacy upon events.

Instead, attention has been directed towards the positive arguments that Anomalous Monism offers in defending the causal efficacy of the mind. There are three important ways that the mental can be regarded as making a significant difference, so far as Davidson is concerned; these are presented in the next three Subsections. Firstly, he believes that his Extensionalist view of events avoids the problem of attributing causal efficacy to mental properties. Secondly, the supervenience relation is a determination relation between mental and physical, by which mental properties have an effect on physical properties. Thirdly, there exist non-strict laws that hold between the mental and the physical, which give a perfectly intelligible justification for attributing causal powers to mental properties. To each of these explanations, critics have found worrying problems. These strongly suggest that Anomalous Monism, as so far developed, has failed to defend itself against the threat of Epiphenomenalism. Finally, Subsection Five remarks upon the oddity of there being no account of how our sensations possess their particular characters, even though this is a salient case of the body affecting the mind.

### Subsection 1 : Does Anomalous Monism entail 'Type Epiphenomenalism'?

It is important to identify two distinct ways in which mental properties may be thus considered redundant, following a distinction made by C.D. Broad in "Mind and its Place in Nature". He noticed that the Epiphenomenalist position is subtly different depending upon whether events are considered to have both mental and physical properties, or whether events are thought of as either solely mental or solely physical. In the first case, the mind is Epiphenomenal when an event causes another purely because of its physical properties and never because of its mental properties. In the second case, the mind is Epiphenomenal in the sense that a mental event is caused by a physical event, but a mental event causes neither sort of event whatsoever. These distinctions have become referred to as respectively 'Type' and 'Token' Epiphenomenalism.

One characterisation formulates them thus:

*Type Epiphenomenalism* (Type-E). (a) Events can be causes in virtue of falling under physical types, but (b) events cannot be causes in virtue of falling under mental types.

*Token Epiphenomenalism* (Token-E). (i) Physical events can cause mental events, but (ii) mental events cannot cause anything.<sup>11</sup>

Critics are happy to accept that an event's having mental properties will make a difference to the causal relations in which it is involved, for this follows from the supervenient relationship between the mental and the physical. This holds that two events cannot have the same physical description, and yet one event have some mental property the other event lacks. Further, if events are individuated one from another according to the descriptions that are true of them, then in this sense, mental properties are as definitive of that event as are its physical ones. As has been accepted, the mental properties of an event are relevant to the events that succeed it, so the problem does not lie with Token Epiphenomenalism.

Rather, critics are concerned that reference to the mental properties of an event serves merely to help individuate that event amongst others, but that mental properties have no causal significance to the actual sequence of one event following another. The debate centres upon whether Anomalous Monism entails Type Epiphenomenalism, or whether it is merely consistent with it. The serious charge of entailment is examined and rejected by Brian McLaughlin in an article titled "Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical". In broad outline, his argument runs as follows.

The critics who press the charge of Type-E, assert that the mental properties of an event are causally inert. Whenever an event  $e_1$  is said to be the cause of another event  $e_2$ , it cannot be the case that  $e_2$  was brought about in virtue of any of the mental properties of  $e_1$ , even though  $e_1$  may with some propriety be referred to in terms of its mental properties. The problem for supporters of Anomalous Monism, is how to hold both the principle that there are neither strict psychological nor psycho-physical laws, as well as the principle that if events are related, then they fall under a strict law. Their critics argue that such a commitment implies that Type-E is true.

McLaughlin identifies a 'family' of arguments in this vein:

The leading argument that emerges here is this: The Principle of Causality implies that only nomic properties are causal. Anomalism implies that no mental properties are nomic. So Anomalism and the Principle of Causality imply that no mental properties are causal, that is, that mental properties are Epiphenomenal.<sup>12</sup>

---

<sup>11</sup> McLaughlin 1989. Pp 109-110.

<sup>12</sup> McLaughlin 1989. P 120.

One thing that each member of this 'family' share, is the blurring of the difference between 'strict' and 'non-strict' laws. Anomalous Monism is committed to the assumption that the psychological and psycho-physical connections cannot be formulated into strict laws. This, however, is not to say that these connections cannot similarly be formulated into non-strict laws; in fact, Anomalous Monism assumes that there are generalisations that relate events at the mental level, as well as generalisations that relate events between the mental and physical levels. What emerges from the above distinction, is that critics tacitly assume that non-strict laws cannot be causal laws. Here a 'causal law' is defined as one such that, if events fall within the scope of that law, then those events count as being causally related. McLaughlin recasts the critics argument, to incorporate the above distinction:

- i) The Principle of Causality implies that properties related according to strict laws are the only causal properties.
  - ii) Anomalism implies that mental properties are only related according to non-strict laws.
- Therefore:
- iii) i and ii together imply that no mental property is a causal property.

The truth of premise i) is challenged by drawing attention to a move from what is *a* logically necessary property of a causal event, to that being the *only* logically necessary property. However, McLaughlin is able to provide a wide range of instances, which include metaphysical as well as logical necessity, where this inference is not justified. Here are just two of those examples. An object has a certain temperature *T* in virtue of having a certain mean kinetic energy *E*. However, *E* might be multi-realizable at the micro-level, including being realized by the property *m(E)*. Thus some object might have temperature *T* in virtue of having property *m(E)*, where that property is not identical with *E*. Again, one can be a brother in virtue of being a male sibling, or in virtue of being a male with a sister. These properties, though not identical, are both logically necessary to qualify an object as being a brother; but the fact that these properties are logically necessary, does not entail that they are the only such properties.

He concludes that the Principle of Causation does not imply that only those properties that can be related by a strict law are causally efficacious. Rather, it implies that if one event is related to another in virtue of some causation grounding relationship, then they are causally related - and if they are causally related, then they fall under some strict law. The foregoing argument shows that it does not follow that only strict laws are causation grounding. In this way, critics of Anomalous Monism underestimate the range of properties that might be causation grounding.

As McLaughlin points out, for all that has been said:

...if an event participates in a causal relation in virtue of having a certain property, then the event participates in the relation in virtue of having some nomic property or other. And this leaves open whether an event can participate in a causal relation in virtue of having a non-nomic property.<sup>13</sup>

Subsection 2 : Extensionalism as a defence against Epiphenomenalism.

Against the charge that Anomalous Monism provides no place for the mind to have an effect upon the world, Davidson responds by pointing out that with his particular view of events, it makes no literal sense to talk of events being causes in virtue of their properties. Thus our clamouring for some tangible causal property is no more than a reversion to a distinct and faulty view of causation. Events follow one another in a sequence. Each event may be referred to according to whichever properties it possesses. However, alluding to the properties of the events to explain their sequence, is to confuse 'causal explanation' with 'causation'; Event  $e_1$  is followed by event  $e_2$  no matter how described. A causal explanation for this sequence must advert to the physical properties, but that does not mean that the mental properties of  $e_1$  have no influence upon  $e_2$ . All that is implied is that mental properties have no place in a genuine causal explanation, though there is nothing improper about their being adverted to in generalisations accounting for why such a sequence occurred. Accordingly, Davidson feels that he is not obliged to provide a reply to his critics; rather, they should attend to the implications of the Extensionalist view of events.

In his article "On Davidson's Response to the Charge of Epiphenomenalism", McLaughlin argues that alluding to the Extensional nature of events does not excuse Davidson from having to answer his critics. McLaughlin points to certain relations, such as 'heavier than' and 'higher than' that are also Extensionalist, and yet this is quite compatible with their being relations in virtue of the properties of the entities related:

That causal relations are extensional relations between events is straightforwardly compatible with the claim that when events are causally related, they are so in virtue of something about each. Indeed, typically, when a particular bears an extensional relation to another particular, the particulars are so related in virtue of something about each.<sup>14</sup>

One might be led to deny properties a rôle in determining relations, through ignoring an ambiguity in the way a property might be thought to play such a rôle. To illustrate, consider once more the case of the stone in the tyre. Because of the stone's hardness, every time it hit the road's surface, it made a tapping sound. More generally, an event 'c' (the stone hitting the road), causes event 'e'

---

<sup>13</sup> McLaughlin 1989. P 127.

<sup>14</sup> McLaughlin 1993. P 32.

(the tapping sound), in virtue of 'c's having the property F (hardness). On one construal, this is taken to imply that 'c's having F caused 'e'. Which is to say, that a property F, rather than an event, caused 'e'. Davidson denies this construal, saying that 'c' was the cause of 'e'. However this is only one interpretation of the relation between F and 'e'. McLaughlin's point is that this disregards the nature of 'e', particularly the relation between F and a property G of 'e'.

By analogy, consider the weighs-less-than relation that holds between two entities 'a' and 'b'. There are a number of properties of 'a' and 'b', including their weighing respectively ten and twelve kilos. If we say that 'a' weighs less than 'b', then this is true regardless of how we chose to describe these entities. That is compatible with Davidson's position. However, in order for it to be true that the entities stand in that relation, they must have some properties that make their weights different. Further, it must be that the property of 'a' is such that it is smaller than 'b' in comparison. Here their properties clearly make a difference, but not the specific properties of weighing ten or twelve kilos, for instance. The relation 'weighs-less-than' picks out the property type that is going to be relevant in sorting the entities. It also specifies what it is about the tokening of that type which is of interest. So the relation 'taller-than' picks out height that is the property under consideration, and specifies that the relation sorts entities according to how they token that property - how tall they are. On its own the property of being four foot tall, does not of itself determine anything; only in the context of the relation taller-than, and any other entity with which it is being compared, does that particular property become significant. Once entities have been sorted according to the relation, then it is irrelevant how they are referenced. Similarly with the causation relation; an event 'C' having a certain property becomes relevant to event 'D' when they are causally related and 'C' has properties that are suitably compatible with the properties of 'D'.

One should briefly consider another difficulty that faces an Extensional view of events, when it is employed to relate mental states. According to this view, events are related to one another, no matter how those events are described - provided that the description is true. However, the intentional nature of our thoughts makes it impossible to guarantee that one can classify their contents regardless of how they are described. For instance, I am sitting a multi-choice exam on General Knowledge. One question asks: "When did the author of 'Down and Out in Paris and London' die?" Since I happen to know that Orwell died in 1949, I duly tick that box. Another question asks: "When did the author Eric Blair die?". Lacking the knowledge that 'George Orwell' was the pseudonym of Eric Blair, I tick a box indicating that I do not know. The problem is, the first action is truly described as answering in virtue of knowing when Eric Blair died, since he was the author in question. Yet, this is a contradiction of the second action, which indicates that I did not know when Blair died. Whatever the case for the

relation between physical properties, the relation between mental properties does not fit happily within an Extensionalist conception of events.

### Subsection 3 : Supervenience as a defence against Epiphenomenalism.

There are two separate worries about supervenience; whether the relation is compatible with Anomalous Monism, and whether it is possible to make any useful appeal to this relation which illuminates the way the mind and the body influence each other. The second worry will be addressed in the next section. Here the issue is whether one can endorse supervenience whilst denying the existence of psycho-physical laws. Jaegwon Kim has argued that either the notion of supervenience is too weak to be of interest to a monist, or it is a notion which ultimately presupposes the existence of law-like regularities that connect the mental with the physical. If he is correct, then the thesis that the mental supervenes upon the physical is inconsistent with the principles of Anomalous Monism. In which case, supporters of this position are disqualified from employing supervenience, as a means of demonstrating how the mind and body are capable of interacting.

Whilst there is considerable debate concerning which formulation of supervenience is most suitable for the purposes of Anomalous Monism, including an argument of Horgan suggesting that the difference between 'Weak' and 'Strong' is merely a red herring, I intend to concentrate on the formulation used in Kim's replies, which appears to be accepted by Davidson himself. To put it informally, supervenience is taken to imply that a change in mental properties is always accompanied by a change in physical properties, and vice versa.

It should be noted that, although a change in a mental property 'M' of an event 'E' is accompanied by change in the physical properties 'P' of 'E', this is not intended to imply that a change in 'M' will always be accompanied by the same change in 'P' on all occasions. For, if this were so, then there would be a danger that the anomalous relations between mental properties would reflect the nomological relations between physical properties, in a manner that is incompatible with Anomalous Monism. The assumption that a mental property can be tokened by an event without there being any implication for the physical properties tokened by that event, reflects the idea that the mind is physically multi-realizable.

Kim is ready to accept this formulation. However, it only treats of mental-physical supervenience. Since no understanding of the mind-body relation would be complete without an account of how the physical affects the mental, Davidson is obliged to show that supervenience here too does not imply the existence of psycho-physical laws; and this, Kim alleges, he cannot do. Rather:

...mind-body supervenience...can be explained in terms of the existence of generalizations from the subvenient to the supervenient, thus: whenever

anything has mental property *M* there is some physical property *Q* such that it has *Q* and everything that has *Q* has *M*.<sup>15</sup>

We can accept that the above formulation does not imply 'like mental properties supervene on like physical properties'. However, identical physical properties will always be associated with identical mental properties. If this is so, then the law-like manner in which physical properties are related, will be mirrored in the sequence of mental properties. Not only does this contradict the thesis of the anomaly of the mental, but it also makes it harder to see what difference the mental properties make in the way events are ordered.

The following is an attempt to make it clearer how relations between the mind and the body are problematic for Anomalous Monism. Consider a series of events,  $E_1$  to  $E_n$ . The physical properties of these events are related according to strict laws; the mental properties of these events, on the other hand, are related according to constraints of rationality. It is assumed that these accounts are independent of each other. The problem of mental causation, is how to account for the possibility of mental properties affecting physical properties - and vice versa - within the sequence  $E_1$  to  $E_n$ ? Intuitively, we need to make sense of the idea that the sequence is so ordered, not merely because of the physical and rational constraints, but also according to some principled way the mind and the body interact; in other words, that there are three separate influences on events - physical laws, rational principles, and mind-body relations.

Anomalous Monism makes use of the supervenience relation, to illustrate how the mind interacts with the body. According to the supervenience thesis, the mental properties determine the physical properties of an event, at least to the extent that there can be no change in the mental without a change in the physical; nor can there be a change in the physical without a change in the mental. There is a problem of how 'determine' is meant to be understood - at least two uses spring to mind. In one sense, to determine, is to necessitate; a particular outcome, 'x', is sometimes said to be determined by the prior circumstances, 'y', implying that granted 'y' obtained, it was inevitable that 'x' should occur. The other sense is much looser: to determine is merely to stipulate how to select from a range of possibilities. For instance, some goals determine a range of suitable behaviour, without determining the manner in which the goal is achieved. If the mental determines the physical in the first sense, then changes at a physical level should be understood in terms of mental properties, and not according to the axioms of Physics. Further, the idea that the physical is thus subordinate to the mental, contradicts our firm belief that the physical is capable of affecting the mind. Because of these difficulties, the second sense of 'determine' is going to be adopted in the remainder of the enquiry.

---

<sup>15</sup> Kim. 1993. P 23.



There are three possible combinations of mental-physical property relations which are worth considering; where two mental properties are related, where the mental affects the physical, and where the physical affects the mental.

First the case where mental properties are said to affect each other, such as when the belief that it is sunny, brings about a desire to abandon work in favour of a sleep outside. There will be an holistic explanation for the sequence, belief  $M_1$  resulting in the desire  $M_2$ . It will be matched by a description of those events in physical terms,  $P_1$  and  $P_2$ , whose sequence will be explained according to strict laws. Here there is a problem as to how  $P_2$  came into existence, since there are two rival claims; one that it is the subvenient base of  $M_2$ , the other that it is the nomological consequence of  $P_1$ .

It is equally problematic to see how the mental affects the physical. Take as an example, the desire to gain someone's attention by shouting at them. Again there will be a mental property corresponding with the desire  $M_1$ , which will supervene on a physical property  $P_1$ , which in turn will cause the shouting, event  $P_2$ . Once more, there is a competition for determining how to account for the shouting; that is, whether  $P_2$  is the nomological consequence of  $P_1$ , or whether it is the subvenient base of  $M_2$ .

Not surprisingly, this problem arises in cases of the physical affecting the mental, such as when the olfactory properties of gun powder remind me of my adventurous youth. In this instance, the physical property of the smell,  $P_1$ , will cause  $P_2$ , which in turn will be the subvenient base for  $M_2$  - the memory of those crazy days in the West (Presumably,  $P_1$  will be accompanied by a sensation  $M_1$ , which is what brings about the recollection.). This gives rise to the problem of having to adjudicate between the appropriate source of the memories - was it a consequence of the relation between  $P_1$  and  $P_2$ , or  $M_1$  and  $M_2$ ?

In each case, both sets of properties are related according to what are essentially different constraints - either by physical law, or by rationality. Further, the physical properties are both determined by a supervenience relation, as well as a nomological relation. The pattern of physical events as described by strict laws, will match the pattern of mental events according to supervenience. Therefore, though the supervenience thesis carries no implication for the law-like association of physical properties with mental properties, the nomological relation of physical properties will carry across to the relations between mental properties. If it is a law that 'a's follow 'b's, then the rôle of supervenience in determining the properties of an event must carry the responsibility for those properties exhibiting a law-like pattern: hence, the holistic nature of mind is law-like in a physical sense. This consequence, is in conflict with the third of Davidson's principles - the possibility of accounting for these regularities between the occurrence of mental and physical properties, is ruled out by Principle three, 'the Anomalism of the mental'.

There appears to be no way of settling the dispute concerning why any particular property is instantiated by any particular event. As Kim remarks, for all that is stipulated by the supervenience thesis, there is some possible world where the sequence of physical events is identical with those in our world (that is, a comprehensive causal explanation may be given for that sequence), and yet the mental properties of those events could be completely absent. If that is a possibility, then it is not clear that Supervenience implies that the mental has an influence upon the physical.

In order to justify our belief that the mind is able both to influence and be influenced by the body, we require some theory of property relations to this effect, that is compatible with our theories concerning nomological and holistic relations. The problem with supervenience, is that it leads to a dilemma; either mental properties become law-like, which is incompatible with their intentional nature; or they become Epiphenomenal, against our firmest beliefs to the contrary. Either way, neither outcome is acceptable to supporters of Anomalous Monism. It finally remains to be seen whether the notion of non-strict laws can be employed to elucidate the nature of the relation between mind and body.

#### Subsection 4 : Non-strict laws as a defence against Epiphenomenalism.

Davidson points to a similarity between the content of psychology, and that of the 'special sciences', which contain generalisations and 'ceteris paribus' qualifications. No-one would withhold the scientific status from these disciplines, simply because our present understanding of their subject matter prevents us from formulating strict laws similar to those of physics. Of itself, being the subject of a discipline that accepts generalisations, should not disqualify it from being involved in causal relations; what is good enough for geology, is good enough for psychology. Whilst only physical properties may legitimately be cited as causes in strict laws, there is nothing to prevent mental properties in playing a rôle in such generalisations as "Chemical, mechanical or electrical damage to parts of the body will - ceteris paribus - result in the sensation of pain."

One is constantly struck by the similarity between the weather and human nature. Our meteorological knowledge is certainly based upon a less-than perfect understanding of how weather systems work. For instance, we know that when a warm air mass passes over a cold air mass, then the moisture in the atmosphere will condense and form mist. However, we also recognise that there are exceptions to this rule: the precise conditions that need to prevail before mist forms, are not known to us - or ever will be, so far as we can tell. Yet, these gaps do not subvert our faith in the accuracy of weather forecasts (at least, no more than is consistent with an understandable cultural scepticism).

Davidson cites Fodor approvingly, for coming up with arguments that support this contention. In "Making Mind Matter

More”, Fodor notes that if you limit causal efficacy only to those properties that belong to the vocabulary of physics, then you will be denying the truthfulness of all the special sciences. He is careful to point out that the ‘intentionality’ of psychology, is not what separates it from physics. Rather, it is the failure to establish the precise conditions under which certain events take place. One difference between strict and non-strict laws, emerges with respect to how one event brings about another. When the law is strict, then there is no answer to this question - the link between the events, is in some sense metaphysically basic. So far as non-strict laws are concerned, one event brings about another according to some mechanism. The broadness of the generalisation, will depend upon how complete is our understanding of this mechanism. It would be an error to think that a gap in complete knowledge implies complete ignorance. For this reason, it would simply be snobbery to refrain from according causal significance to properties that find no place in physics text books. Therefore, insofar as mental properties appear in generalisations, then their causal powers are just as respectable as those of physics; and insofar as they have these powers, it makes sense to say that they ‘make a difference’.

Kim makes a series of points against the acceptance of non-strict laws by an Anomalous Monist. Firstly, he cautions against the embracing the idea that Fodor has of ‘non-strictness’, since lying behind Fodor’s thinking, lies a faith in the eventual refinement of every psycho-physical law such that eventually they do not need to be tightened-up.

Indeed, Fodor writes:

“Eventually” means: Either the law is implemented by a physical mechanism, or its implementation depends on a lower-level law which is itself either implemented by a physical mechanism or is dependent on a still lower-level law which is itself either implemented by a physical mechanism or...etc. Since only finite chains of implementation are allowed, you have to get to a physical mechanism “eventually”.<sup>16</sup>

This is simply to allow the possibility of strict psycho-physical laws, contradicting the principle of the Anomalism of the mental. So, if Davidson is thinking of non-strict laws, they have to be quite different from the sort Fodor has in mind.

The natural consequence of the above objection, is to examine whether any construal of non-strictness is compatible with the spirit of the anomaly of the mental. One of the virtues of this position is that it recognises that the relations between intentional mental events are essentially different from those relating physical states. Accordingly, it rules out the possibility of laws relating the mental with the physical, in virtue of the fact that they would be attempting to relate two categorically separate parts.

---

<sup>16</sup> Fodor. 1990. P 159.

Kim points out:

...if each domain is to retain its own integrity, there cannot be laws connecting them. It isn't clear why this argument, if it succeeds in banning strict psycho-physical laws, doesn't banish non-strict ones as well.<sup>17</sup>

Moreover, alluding to the similarity between psychology and other special sciences like geology, only shows that explanatory truth is *compatible* with non-strict laws. It does not guarantee that such laws will be true. Whatever the relation between the strict and non-strict laws, certainly in the examples given, there is a feeling that what relates the vocabulary of geology to that of physics, is importantly different from that between psychology and physics. Such an association appears to have mislaid the point that was made about the relation between mental properties - that they are holistic in character. So far as the formulation of laws are concerned, the predicates of mind and matter simply are not made for each other.

Subsection 5 : Has enough been said about the character of sensations?

Before considering the issue of supervenience further, it might be worth pointing out that there is a preponderance of examples of the mind affecting the body, rather than on the body affecting the mind. One reason for this, is Davidson's work on construing 'reasons' as causes of action; thus, the preoccupation is with a mental-physical relation. However, this tends to set the enquiry off-balance; whilst it may be illuminating to think of 'reasons' as providing a medium for mind-body interaction, it is by no means clear how the body is able, in like fashion, to affect the mind. By neglecting such mental states as sensations, where the body would appear to be affecting the mind, Anomalous Monism is always going to be hard pressed to provide a full account of the relations between the mental and the physical. Invariably, physico-psychological relations will be left unexplained, except for the expectation that they will follow analogously to psycho-physical relations. Until more is said upon this matter, it has to be regarded as a deficiency of Anomalous Monism, that it says so little about the way nature affects our mental states.

### Section Three

Subsection 1 : General problems with Supervenience.

The idea of supervenience strikes some as being more trouble than it is worth. It may be true that you can relate two seemingly disparate states or properties, by observing that you do not get a change in one level without a corresponding change in the other. Yet it is sometimes difficult to get clear what philosophic work has been done by this observation. Discerning this relation does not of itself

---

<sup>17</sup> Kim. 1993. P 25.

answer questions about the ontological status of the levels being related; just because two properties are thus related, carries no implication for whether or not they genuinely exist. For instance, it is plausible that one could relate the properties of superstition with physical properties; the higher level properties of fortune could be thought of as determining the physical events that affect people's lives. In an analogous way to mental supervenience, any change in a person's fortune would have to be reflected in a change in their physical circumstances. Similarly, if the physical circumstances do not alter, then neither will their state of fortune. Doubtless there will be refinements to this facetious use of supervenience, but that only reflects the refinements to supervenience that presently take place in the philosophy of mind. For instance, one can choose from weak, strong, local, regional or global supervenience, according to your fancy.

Inconsistencies arise from the fact that there is no agreed use of the term. An illustration of this may be found in Davidson's use of the supervenient relation. I am thinking here of the introduction into philosophic literature of that incredible creature "Swampman", in the article "Knowing One's Own Mind". 'Swampman' features in a debate about Externalism, in a thought experiment that runs somewhat as follows. Suppose you are walking beside a swamp in a thunder storm. You are remembering the last time you were caught in such a storm, when suddenly a bolt of lightning strikes, and reduces you to a pile of smouldering ash. Simultaneously, the energy from the blast composes a replica of yourself from swampy material. The replica's physical states are exactly so configured as to match those that you would have had were you not so cruelly struck down. The purpose of the experiment is to provide an instance of two physically identical entities that have different mental states:

My replica can't recognize my friends; it can't *recognize* anything, since it never cognized anything in the first place...It can't mean what I do by the word 'house', for example, since the sound 'house' it makes was not learned in a context that would give it the right meaning - or any meaning at all. Indeed, I don't see how my replica can be said to mean anything by the sounds it makes, nor to have any thoughts.<sup>18</sup>

Disregarding the cogency of this argument for the moment<sup>19</sup>, it clearly contradicts one formulation of the supervenience relation given earlier - that there cannot be two events alike in all physical respects, but differing in some mental respect<sup>20</sup>. The Swampman is supposed to be physically identical, yet instead of remembering the last time he was in a thunderstorm, it is argued that it is nonsense to speak of the Swampman remembering anything, since its recent nativity hardly affords such a luxury. Though it is possible to reformulate the supervenience thesis to accommodate Externalism, there is a suspicion that instead of being the bedrock of a theory,

---

<sup>18</sup> Davidson. 1994. P 47.

<sup>19</sup> The case of Swampman is discussed in more detail in Chapter 4, Pp 78-79.

<sup>20</sup> See quote above, P 17.

supervenience is merely a shifting sand of monistic wish-fulfilment. That is, instead of placing constraints upon the positions we are able to adopt, a supervenience relation may be Gerrymandered according to the view we have already chosen.

### Subsection 2 : Horgan's remarks upon Supervenience.

The literature upon Supervenience relations is too vast to do justice in such a small space. A comprehensive survey may be found in an article by Terence Horgan, "From Supervenience to Superdupervenience: Meeting the Demands of a Material World." Horgan raises a number of issues concerning this relation. He feels these have not been fully appreciated by those philosophers who hope that supervenience will resolve all the problems that hitherto beset a Materialist theory of the mind. He immediately gestures toward what will be his primary concern:

The principle moral will be that supervenience relations, in order to figure in a broadly materialistic worldview, must be explainable rather than *sui generis*.<sup>21</sup>

It is Horgan's main worry that when one examines the past use of the idea of a supervenience relation, it becomes clear that such usage was not compatible with a Materialist cast of mind. He traces its origins to the British Emergentism, in the early Twentieth Century, where the notion was introduced to relate higher-level emergent properties to lower level physical properties. However, this was a relation where the higher-level properties had causal powers that were novel, in the sense that they were not taken to be mere compounds of pre-emergent physical properties. Similarly, the nature of the connection was something for which no explanation could be given, but which had to be taken as something metaphysically fundamental. Clearly, this analysis of supervenience, is unavailable to the Materialist. No Materialist could countenance having to disregard a belief in the completeness of Physics, nor having to accept an inter-level connection whose understanding lay outside the physical domain.

Things get no better with the next application of this relation. In "Principia Ethica", G.E.Moore embraced the idea to connect non-natural properties of value, with the natural properties of descriptions. Here again, there are two levels of discourse; the higher-level moral properties, and the lower-level physical ones. Moral properties supervened upon natural properties, such that if two objects were identical in their physical state, then they cannot differ in their moral state. Sadly, the moral properties being thus related, were regarded as non-natural; so again, this usage is one that has no Materialistic respectability.

'Supervenience', as a philosophic expression, first became used by R.M.Hare in "The Language of Morals". So far as Hare was concerned, the relation illustrated a constraint upon how moral

---

<sup>21</sup> Horgan. 1993. P 556.

concepts could be used in Moral discourse. There was no intention that the relation should be between putative moral facts or properties. It was simply to show how moral terms commended or proscribed, as related to physical properties in the world. Again, there were two levels of discourse, such that a change in one level would necessitate a change in the other; but once more, there was nothing in this conception of supervenience that could be utilised by Materialism.

The object of this overview, is to raise the following point. The potential for a supervenience relation between two levels of discourse does not entail either of the following claims; that the relation concerns 'properties', rather than 'social practices'; or that where the relation does concern properties, that the relation holds between properties of a single ontological kind. Of itself, the existence of a supervenience relation is no guarantee of the existence of the supervening terms in a Realist sense. Further, the fact that the relation can be understood in terms compatible with a Materialist viewpoint, as with Hare's usage, does not mean that the relation between the higher and lower levels would be one acceptable to a Materialist. If supervenience is to play a rôle in a monist theory of mind, then the relation must be seen as being between two entities of the same ontological type, and must exhibit a nature that is recognisably part of the world. To mark how distinctive this requirement is, such a relation is dubbed 'Superdupervenience'. Horgan observes that:

...any genuinely materialistic metaphysics should countenance inter-level supervenience connections only if they are explainable in a materialistically acceptable way, and should countenance *ontological* inter-level supervenience relations only if they are *robustly* explainable in a materialistically acceptable way.<sup>22</sup>

Horgan addresses some remarks to those who hope to establish such connections. He suggests that they require a set of facts,  $F_1$  to  $F_n$ , which could combine with physical facts to connect the lower level with the higher level mental properties. However, it should be possible to have some understanding of how  $F_1$  to  $F_n$ , rather than any other facts, constitute legitimate connections between high and low level properties. This in turn requires a comprehension of the nature of this connection, such as would enable the selection of just those facts which were illuminating, from all the facts available. Finally, there must be some justification for thinking that these factual resources actually exist.

Even were it possible to find a satisfactory means of meeting these requirements, there is a further concern that must be addressed. Horgan notes that both Morality and Intentionality share the feature of having standards of correctness. Just as it is possible to act rightly or wrongly, in the ethical sense, so too is it possible to act rationally or irrationally. If there is to be a physical account of behaviour, then there must be a physical correlate to this

---

<sup>22</sup> Horgan. 1993. P 563.

property of 'rationality'. Yet the 'is-ought' gap that bedevilled attempts to relate moral properties to natural ones, re-appears when one attempts to relate other normative properties with natural ones. Horgan concludes that:

Explaining ontological supervenience relations in a materialistically acceptable way looks like being a very daunting task, whose difficulty suggests the need for materialists to consider seriously the prospects for preservative irrealism about much of our higher-order discourse.<sup>23</sup>

### Subsection 3 : Wilson's amendment to 'Superdupervenience'.

Wilson questions the suitability of 'materialistically acceptable explanations' (hereafter indicated as 'explanation<sup>s</sup>') as criteria by which to judge whether or not a supervenience relation is 'Superduper'. She argues that any supervenience relation that satisfies these criteria should not automatically be thought of as carrying the Materialist stamp of approval. The fact that A provides an explanation<sup>s</sup> for B, is neither necessary nor sufficient for the relation between A and B being appropriate to a Materialist metaphysics.

It is not sufficient, since one can envisage cases where the explanation of B in terms of A was suitable, but where that explanation is in competition with others. The problem is not merely to provide an explanation of the appropriate type, but also to be able to determine that it is *the* correct explanation. Wilson employs Horgan's 'liquidity' example to illustrate this point. It is possible to provide an explanation<sup>s</sup> for the higher-level features of liquidity, in terms of the behaviour of substances with certain low-level structures. With this explanation, the features that define liquidity are regarded as being nothing over and above the properties of molecules relative to a particular context. However, there is nothing to rule out an alternative explanation for such higher-level features - that they are *caused* by the low-level molecular properties. Where this is the case, it is plausible that there may be novel, or emergent, causal powers associated with liquids, which cannot be explained in terms of the low-level molecular properties that brought it into being. Such an explanation would not be acceptable to Materialism. Therefore, the existence of an explanation<sup>s</sup> of B in terms of A, is not sufficient to guarantee that the supervenience relation between A and B is Superduper in the required sense. All it does, is guarantee that the proffered supervenience relation is intelligible.

There are a number of good reasons for thinking that the explanation<sup>s</sup> condition is not even a necessary feature of a Superduper relation. The first thing to note is that an epistemic condition - 'fitness of explanation' - is being used to perform a metaphysical task - 'classifying ontological relations'. Certainly, where one is able to explain A in terms of B, then there is a metaphysical relation between A and B. This only sanctions the use of explanations as indicators of where such metaphysical relations

---

<sup>23</sup> Horgan. 1993. Pp 581-2.



exist. Wilson feels that there are more appropriate markers of such relations - 'causation' being preferable to 'explanation', in its having a distinctly metaphysical character.

There is also an incompatibility between the Materialist's belief in metaphysical unity, and the variegation evinced by our explanations of reality. It has been strongly suggested that the findings of the different scientific disciplines cannot be reduced to one unified science<sup>24</sup>. If 'explanations' do not reflect the Materialist's idea of the unity of reality, then their suitability as arbiters of metaphysical relations, has to be questioned.

Finally, granted the limits of our cognitive capacities, it is unlikely that we will be able to explain everything. There may well be potential explanations which we are not in a position to articulate. In this sense, explanation is seen as being an inadequate indicator of metaphysical relations, insofar there is no assurance that once everything is explained, all relations will be revealed. Importantly, although an explanation of A in terms of B, entails a relation between them, it is not clear that the entailment goes the other way round. It is not clear that the relation between A and B will always be such as to admit of an explanation of B in terms of A.

Wilson notes that the issue between Emergentists and Materialists, concerns the origins of the causal powers of properties. The Emergentists allow for the creation of such powers over and above those of their lower-level properties. The Materialists deny that there should be any powers except those revealed in lower-level properties. As an alternative to explanation<sup>s</sup>, Wilson suggests that what is definitive of Superduper relations, is the following constraint - referred to as the Condition on Causal Powers (or CCP):

Each individual causal power associated with a supervenient property is numerically identical with a causal power associated with its base property.<sup>25</sup>

There are a number of advantages to employing this condition. It is capable of ruling out non-Materialist uses of Supervenience. It allows for the distinctiveness of the higher-level properties, by regarding them as being a proper subset of the low-level properties. It allows for the multi-realizability of higher-level states in terms of lower-level ones: that is, there may be more than one set of lower-level properties to which the higher-level subset could belong. Also, it rules out the possibility of over-determination, since the higher-level causal properties cannot be seen as being in competition with the lower-level ones. Intuitively, the proper subset of causal powers may be regarded as a specific concatenation of members of the set of powers classified at the lower-level. It is not the case that a different set of powers is in operation.

---

<sup>24</sup> For instance, see Cartwright 1996.

<sup>25</sup> Wilson. 1999. P 42.

Wilson goes on to discuss three non-reductive versions of accounting for the mind body relation, in which this condition is satisfied, and concludes that:

Arriving at a physicalistically acceptable supervenience relation has...less to do with an epistemic requirement on explanation than with a metaphysical requirement on causal powers.<sup>26</sup>

There are two drawbacks to the above amendment. The first is that, although it brings the requirement for a Superduper relation into sharper focus than Horgan, it still remains doubtful whether such a relation exists between mental and physical states. Admittedly, we now have a clearer idea of when such a relation is present, but only in virtue of explaining how the causal properties of the mind can be identified as being a proper subset of those of the body. This is a problem that supervenience was meant to solve. Further, it just is not clear that the distinctiveness of the mind is accounted for by its being a proper subset of the properties of the brain. The set of Psychological causes does not appear to have *any* relation to those causes we associate with the brain, let alone being a subset. If they are such a subset, then it is not of a set of lower-level causal properties with which we are at present familiar. Since it is unlikely that science will come up with any significant extension to these, it is no more than a blind hope that the supervenient relation between the mind and the body does in fact comply with the 'Condition on Causal Powers'.

The second drawback is that the issue of psycho-physical *causation* is just one aspect of the overall problem facing Materialists. Of primary importance, is the identity relation. It is not simply a matter of wondering why physical states are capable of causing mental states, it is also unclear how a mental state can be identical with a physical state. That is, how something we know by its mental properties could at the same time be known through its physical properties, when the inter-relations that these properties have within their respective domains are so distinct. The expectation is that by employing the CCP we may eventually discover the Superdupervenience relation. To do this, we must have first identified the causal powers of the mind as belonging to a set of causal powers of the body. However, since the supervenience relation was originally employed to make this identification between mind and body, and since identification is a prerequisite for judging the propriety of supervenience relations, then it looks as if the application of CCP is impossible.

#### Subsection 4 : Conclusion.

None of this implies that the mental does not supervene on the physical; rather, the implication is that the relation of supervenience may be employed by philosophers of mind of all persuasions. In effect, Anomalous Monism is the first thesis to

---

<sup>26</sup> Wilson. 1999. P 51.

employ the concept to relate properties of the same ontological kind. If Anomalous Monism is to refute the charge of Epiphenomenalism by adverting to the supervenience relation between mental and physical properties, then it has to satisfy all the doubts raised above; so far, all that is established, is that the mental and the physical *could* be so related, but that is a long way from showing that in fact, that *is* how they are related. For the moment, one must conclude that the threat of Epiphenomenalism is not averted, simply by the Anomalous Monist's faith in supervenience<sup>27</sup>.

#### Section Four

Defenders of Anomalous Monism claim that many criticisms are the product of confusing the terms 'cause' and 'causal explanation'<sup>28</sup>. Typically, when we advert to the causes of a particular event, we do so in virtue of only some of the properties of the event - those that are referred to by some strict law. The fact that mental properties are not cited as causes has nothing to do with whether or not they are influential in bringing about effects: rather it is because they belong to a different nomological category from those of the physical. The reference to 'causes' takes place solely in the context of explanation. Problems arise when identifying the relation that determines a particular sequence of events, with the relations to which we refer in the explanation of that sequence. This section presents two arguments which suggest that the connection between 'cause' and 'causal explanation' is stronger than is portrayed by Davidson's position.

##### Subsection 1 : Crane - some explanations are better than others.

Doubtless we point to mental states in explaining certain events - such as when we explain someone's behaviour in terms of their being 'short-tempered'. What is not so clear, is that one is mistaken in thinking of these states as being 'causes' outwith the activity of explanation. Intuitively, one supposes that one cites such properties as appear in causal explanations, in virtue of those properties being the cause of one event following another. However, Anomalous Monism stipulates that, properly speaking, only physical properties of events are 'causal'. This assumption, embodied in the second Principle, has created a lot of opposition. Tim Crane, in an article titled 'Mental Causation', notes that Davidson is not oblivious to the difference reference to properties makes in our enquiries, but that he treats their rôle somewhat differently from its standard usage:

This is not to say that Davidson cannot give any answer to the question, 'was it in virtue of the brick's mass or its colour that it broke the window?'. But he

---

<sup>27</sup>These points apply to Materialism generally.

<sup>28</sup> See above, P 19.

will regard this question not as a question about the efficacious *properties* of the cause, but as a request for an informative causal *explanation*.<sup>29</sup>

The problem is not whether there is any way for the explanation offered to be informative, which does not advert to properties; but rather, how to individuate the various explanations available, according to how one might be better than another. It seems that one virtue of accepting that causes have effects in virtue of either their mental or physical properties, is that it enabled the distinction between those properties of an event which were relevant and those which were not.

The trouble is that Davidson's theory seems to leave us unable to answer the question of why certain explanations are better than others by invoking the efficacious *features* of reality.<sup>30</sup>

Crane draws attention to the fact that you cannot sever the link between a correct explanation and the properties of the events that are cited in that explanation. If the two are kept distinct, then there appears to be no foundation for preferring one explanation to another.

### Subsection 2 : Sosa - the drawbacks to treating all causes equally.

This line of criticism is taken up by Ernest Sosa in the article: "Mind-Body Interaction and Supervenient Causation". Sosa draws an analogy between the causal relevance of mental properties, as defined by Anomalous Monism, and the causal relevance of the loudness of a shot that kills someone. Given that a person can be killed equally as well with a gun fitted with a silencer as without, then it follows that the sound a gun makes when it is fired need not be a determining factor in whether or not a person dies. Similarly, which mental properties an event has, need make no difference to the effect that is brought about by that event.

Davidson points out that Sosa's example of the gunshot wrongly associates the question of loudness or softness to the same death, whereas according to Anomalous Monists, an event that has a different description brings about a different effect. The sound the shot makes is not to be ignored, since there are two deaths taking place. It is in virtue of one gun being silenced, that the deaths can be individuated. Accordingly, if the sound is to have a rôle here, then by Sosa's analogy, so too will the mind. This reply brings the nature of the objection into sharper focus. Accepting that Anomalous Monism does allow for mental properties to make a difference in this way, suggests that some sort of Essentialism is implicit in the thesis. In 'Davidson's Thinking Causes', Sosa responds by accepting that the particular death is affected by the sound of the shot, but to that extent, it is affected by every

---

<sup>29</sup> Crane. 1995. P 227.

<sup>30</sup> Crane. 1995. P 228.

particular property, mental and physical, that is true of that event. Where each property is considered equally important, it is impossible to discriminate between those properties which serve to individuate the effect, and those which bring the effect into being. The critics suggest, that on this account, the mental properties of an event are allowed no more significance than a speck of dust.

This egalitarian attitude toward properties offends our intuitions in two ways. Firstly, by making each feature of an event equally important, it will be impossible to derive principled generalisations concerning which types of events follow one another; each event would be a unique type. It is supposed that there are salient features of reality by which we can pick out patterns of events; the idea of cause and effect is based upon our ability to determine the existence of these patterns. Secondly, there is a feeling that, although all properties are equal, some are more equal than others: that the muzzle velocity is a far more important factor in explaining why the shot was fatal, rather than that the assassin's thumbprints were whorls rather than loops, or that they preferred tea to coffee.

Sosa notes a further problem with the Anomalous Monist's position upon mental properties. Either they acknowledge their existence, or they deny it. He suggests that if they accept, then the position becomes a form of Property Dualism. It then becomes difficult to see in what sense the position is any longer monistic. There would be an unsolved puzzle concerning how a single substance could be so pliable as to admit to two entirely distinct sets of properties. Furthermore, it is barely coherent to think of a substance whose behaviour is simultaneously subject to two immiscible constraints - those of physical laws, and those of rationality. On the other hand, if they deny the existence of mental properties, then all occurrences of mental states are inexplicable adjuncts to physical states - inexplicable because there are no mental properties to refer to in connection with the accompanying physical state. In both cases, the result is inconsistent with the spirit of Anomalous Monism, either by being dualistic, or by denying that the mental can have an effect upon the physical.

Since time began, Philosophers have been motivated to forego idle pleasures in the search for an answer to the mind-body problem; not surprisingly, they are less than impressed by an account that attributes no more significance to the state of a person's mind than it does to the cleanliness of that person's gun.

### Subsection 3 : Conclusion.

Anomalous Monism promised to respect the irreducibility of mental properties, and in so doing, ensure that the manner in which the mind wielded influence was not derived from its physical nature. Instead, what has emerged from the criticism of this position is that the relation between mental and physical properties will remain problematic whilst it is assumed that they are intrinsically different.

Each attempt to describe how the mind could make a difference has either resulted in consequences contradictory to the principles of Anomalous Monism, or has required a radical overhaul of our everyday intuitions. Rather than accept this dilemma, the intention is to pursue a different line of enquiry. The next two chapters outline the development of Representationalism, and discuss the problems it faces.

### Chapter Three

An alternative strategy, is to place the idea of mental causality firmly at the centre of attention. This chapter examines the family of theories that may loosely be gathered under the heading of 'Functionalism'. D.M. Armstrong offers this description of such a position:

The concept of a mental state is the concept of something that is, characteristically, the cause of certain effects and the effect of certain causes...The effects caused by the mental state will be certain patterns of behaviour of the person in that state...The causes of mental states will be objects and events in the person's environment.<sup>1</sup>

For these theories, the problem of the mind 'not making a difference' was thought to be less acute, since they regard mental states as mediating causally between inputs and outputs. However, not only do they have problems of their own, but the threat of Epiphenomenalism remains. The first Section of the chapter will outline briefly the origins and development of Functionalism, as it has grown in the light of criticisms. As will become evident, it is possible to trace the theory's failings back to the manner in which it evolved. Such problems exemplify the dilemma facing Materialism: whether Mental states may be regarded as knowable from the third person, and thereby qualify as subject matter of a proper science; or whether to regard the workings of the brain as being the key to understanding the mind. This dilemma splits the positions available. The contest between the verificationism of the Behaviourists in one corner, versus the Identity Theorists in the other, manifests itself in the present-day debate between the Externalists and the Individualists<sup>2</sup>.

The second Section treats with two types of objection which face any theory that attempts to place the mind within a natural setting - the possibilities of 'absent' and 'inverted' qualia<sup>3</sup>. As a result of these objections, Functionalism has undergone a number of refinements. These will be discussed in Section Three. Whether or not such changes go far enough, is questioned in Section Four. Here, it is suggested that Functionalism has still left certain questions unanswered. The fifth Section introduces the idea of Representationalism, as a theory that will fill in the deficiencies of the Functionalist account. Not all philosophers accept that we require such a theory; their reasons are briefly presented in Section Six. Finally, since there are two distinct positions that one can take on a Representational theory - referred to as 'Vertical', or 'Horizontal' theories - Section Seven canvasses the arguments for preferring one position to the other. A justification is

---

<sup>1</sup> Armstrong. 1981. Pp 21-22.

<sup>2</sup> This debate will be surveyed in Chapter Four, Section Two, in particular Pp 87-89.

<sup>3</sup> Qualia are taken to be the felt or phenomenal qualities associated with experiences.

offered for thinking that 'Vertical' theories are deeply flawed, and therefore 'Horizontal' theories should be adopted: this position will be the topic of Chapter Four.

### Section One

Present-day Functionalist theories acknowledge their debt to two Materialist accounts of the nature of the Mind, referred to as 'Behaviourism' and the 'Identity Theory' respectively. This Section suggests that Functionalism grew from the attempts to remedy the deficiencies of these theories.

#### Subsection 1 : Behaviourism.

Behaviourism regarded mental states as being dispositions to behave in certain ways, rather than states of some mysterious inner control centre. If psychology were to be a respectable scientific practice, then it should only take as its subject matter, that which was verifiable by observation - the relationship between how an individual was stimulated by their environment, and the responses they made to those stimuli. The correct way of conducting psychological research was to attend to the overt behaviour of individuals, rather than covert relations within the mind. This approach is neatly summed-up by B.F. Skinner:

The practice of looking inside the organism for an explanation of behavior has tended to obscure the variables which are immediately available for a scientific analysis. These variables lie outside the organism, in its immediate environment and in its environmental history. They have a physical status to which the usual techniques of science are adapted, and they make it possible to explain behavior as other subjects are explained in science.<sup>4</sup>

Whilst such a position achieved the desired objective of confuting the fable of the 'Ghost in the machine', it clearly failed to account for the complete range of mental states to be explained. The Behaviourists were accused of having to 'feign anaesthesia' if they were to take seriously the notion that such experiences as joy, sorrow, and pain for instance, only existed as patterns of behaviour. A further embarrassment was the realisation that some 'intelligent' behaviour is exhibited by inanimate machinery. On Behaviourist criteria, such machines would thereby be credited with mental states. So not only did Behaviourism fail to explain how we attribute qualitative mental states to some individuals, it also failed to explain why we withhold such attributions from others.

---

<sup>4</sup> Skinner. 1953. P 31.



Further, as Geach noted<sup>5</sup>, there is a circularity in any analysis of behaviour that involves beliefs and desires. Consider, for example, the case of a person running along a platform beside a stationary train. In order to explain his running, we can suggest that it is because he believes the train is about to depart, and he does not want to miss it. However, this form of explanation is not available to the Behaviourist. On its own, the belief explains nothing; he may entertain the belief, yet that need not be the reason for his running - he might be pursued by a dog. Alternatively, the desire alone explains nothing; he may still want to catch the train, without believing that it is about to depart, but runs merely because he is keen on getting exercise. We can only attribute beliefs to an individual in the case where we are certain what their desires are. However, when we come to analyse those desires, we need to advert to their as yet still unanalysed beliefs. There does not appear to be a way into this circle, and this criticism especially has motivated a search for an alternative to Behaviourism.

### Subsection 2 : Identity theories.

One such contender was the Identity Theory, which identified mental states with brain states. At first glance, the idea is quite promising. Since the mind is commonly regarded as the source of motivation, and since it has become possible to trace the workings of the brain as being the initiating centre of bodily movement, it is natural to identify the workings of one with the other. In this way, the psychological story involving desires and beliefs mediating between sensory information and behaviour, is translated into a neuro-physiological story involving the connections between afferent and efferent neurones by brain states. As J.J.C.Smart puts it:

...in so far as a sensation statement is a report of something, that something is in fact a brain process...When I say that a sensation is a brain process...I am using "is" in the sense of strict identity.<sup>6</sup>

In this manner, beliefs, pains, hopes and so forth will become identified as excitations in certain regions of the central nervous system. One problem was to account for why there was no conceptual identity between the mental state and a brain state. Defenders of the position, such as Smart, likened the identity as being of the same ilk as the identity between lightning and an electrical discharge. These are two descriptions of the same entity, such that being familiar with one description does not mean that one is thereby familiar with the other. Just as it was an empirical discovery that lightning had a straightforward physical composition, so too, we will discover empirically which mental states are identical with which brain states.

---

<sup>5</sup> Geach. 1957. Chapter Four, Pp 7-10.

<sup>6</sup> Smart. 1959. P 145.

Quite apart from the difficulties to be faced in defending such an identity claim, this view shared a further problem of the Behaviourists; it failed to cover all the possible cases where we would be prepared to attribute mentality. However, unlike the Behaviourist, who would be forced to acknowledge mentality in objects where common-sense would withhold it, the Identity Theory is committed to just the opposite - withholding attributions of mentality from creatures where common-sense would extend it. Hilary Putnam described the problem for the identity theorist thus:

He has to specify a physical-chemical state such that *any* organism...is in pain if and only if (a) it possesses a brain of suitable physical-chemical structure; and (b) its brain is in that physical-chemical state. This means that the physical-chemical state in question must be a possible state of a mammalian brain, a reptilian brain, a mollusc's brain...etc. At the same time, it must *not* be a possible (physically possible) state of the brain of any physically possible creature that cannot feel pain. Even if such a state can be found, it must be nomologically certain that it will also be a state of the brain of any extra-terrestrial life that may be found that will be capable of feeling pain before we can even entertain the supposition that it may *be* pain.<sup>7</sup>

An attempt to amend this position, by redefining the relation between mental states and brain states to exist between types of the former and tokens of the latter, only resulted in a very weak form of theory. This version fell prey to the same problem as had Behaviourism, insofar as there seemed to be no restriction on how a mental state might be tokened.

Such arguments as those presented by Putnam, forced philosophers to accept that if we were to attribute mental states to creatures other than ourselves, then we will have to abandon the Identity theory. If pain were to be identified as C-fibre activity, for instance, then one would have to assume that any organism lacking such features, would thereby be unable to feel pain: this seems plainly absurd. Instead, it is assumed that mental states can be realized by a diverse range of physical systems. The idea of multi-realizability becomes very important in assessing where to draw the line between the animate and inanimate. What is common to all realizations of a particular mental state, is that it performs a certain function. For example, what unites my aggressiveness with that of a wasp, is the process by which we are able to act in that particular manner, not the physiology that is in common between us.

Putnam arrived at a development of a theory which appeared to have the virtues of both Behaviourism and the Identity Theory, but which embodied a more appealing principle of how to ascribe mental

---

<sup>7</sup> Putnam. 1967. P 436. See also Fodor 1968, Chapter 3, especially Pp 100-120, for further arguments leading away from type-identity theories toward a Functionalist viewpoint.

states. According to this position, mental states such as pain may be thought of as follows:

...pain is not a brain state, in the sense of a physical-chemical state of the brain (or even the whole nervous system), but another *kind* of state entirely. I propose the hypothesis that pain, or the state of being in pain, is a functional state of a whole organism.<sup>8</sup>

At this point, the idea of Functionalism may be considered to be a serious contender for the job of providing an account of mental states in physical terms. This is not to suggest that the other theories have been absorbed within it; both are being developed by their supporters, independently of efforts to refine Functionalism.

## Section Two

Functionalism quickly became a target for criticism. In two influential papers written by Ned Block - 'Problems with Functionalism' and 'Are Absent Qualia Impossible?' - it became clear that the root objections to Behaviourism and the Identity Theory, carry over to Functionalism. This section describes these objections in the first two Subsections, before outlining the implications upon the viability of Functionalism. Block presents the problem as follows. First he provides a definition of Functionalism:

Functionalism is the doctrine that pain (for example) is identical to a certain functional state, a state definable in terms of its causal relations to inputs, outputs, and other mental states. The functional state with which pain would be identified might be partially characterized in terms of its tendency to be caused by tissue damage, by its tendency to cause the desire to be rid of it, and by its tendency to produce action designed to shield the damaged part of the body from what is taken to cause it.<sup>9</sup>

Block notes that Functionalism has two strands; one which identifies functional states in terms of the a priori concepts used by folk psychology, developed by Lewis<sup>10</sup> and Armstrong for instance; the other, developed by Putnam amongst others, where the functional states are part of empirical psychology. Whereas the former identify the relations between mental states and physical behaviour in terms of common-sense analysis, the latter is a more ambitious project, which uses Turing machine-table states as an analogue of the relation between inputs, processing and outputs. The computer lends itself admirably to the image of multi-realizability, in that a single task can be managed using different processes. Crudely put, the mind is like a

---

<sup>8</sup> Putnam. 1977. P 433.

<sup>9</sup> Block. 1980. P 257.

<sup>10</sup> In 'Mad Pain and Martian Pain', for example. Lewis. 1980.

program that mediates between the sensory inputs and the behavioural outputs.

The objections focus upon two alleged deficiencies of the Functionalist interpretation. The first asks whether qualitative and cognitive states are necessary to the attribution of a mind to an entity; these have become known as absent qualia problems. The second asks whether an entity's being organised in a manner consistent with the above definition, is sufficient for that entity having mental states; these are illustrated by such thought experiments as the 'Chinese Nation'. There is a related argument concerning the possibility of qualitative states being inverted from one individual to another. If this is valid, it indicates that, as a theory, Functionalism is too coarsely grained to account for all varieties of mental state.

### Subsection 1 : Are absent qualia possible?

The absent qualia problem appears most plausible if one is beguiled by the parallels between computers and brains. If the mind is seen as a mediator between sensory input and behavioural output, then it should be possible to tabulate which inputs connect with which outputs. In which case, one can imagine a person whose cranium is populated with tiny people - an Homunculus head - whose task is to carry out instructions according to such a manual. These people receive an input code, which together with their 'present state' code, forms a reference in the manual to an output code. This they key in, and await the next input. Overall, the behaviour of the body they inhabit will be indiscriminable from that of a body using normal mental states to co-ordinate observation and movement. If a Functionalist account of the mind cannot in principle distinguish an object without qualitative experiences, from one where they are present, then it has failed to provide any explanation of what such mental states are. This objection is particularly telling against the Lewisian type of Functionalism, where sameness of mental state is largely determined by common-sense criteria, rather than any specific material composition.

There is a similar objection raised by John Searle<sup>11</sup>, where the focus is upon the failure of Functionalism to account for cognitive states, such as 'believing' and 'understanding'. A person is sat in a room being fed pieces of paper with symbols on them. He matches these symbols with those in a book, in order to find out what sort of symbol he should output. The man never gets to understand that he is being given questions in Chinese, and that he is answering them. He knows neither that he is using Chinese, nor that he is producing responses to questions. All he has, is a correlation between certain forms of input and output. The purpose of this thought experiment is to demonstrate that understanding - semantics - is not equivalent to symbol manipulation according to rules - syntax. Since computers are

---

<sup>11</sup> Searle. 1992. Chapter Two.

syntactically driven, it is argued that like the man in the room, they are unable to comprehend the nature of the tasks they perform. Insofar as implementing a program is not sufficient for understanding, then there must be more to our minds than their simply being programs.

Block makes a similar point against the Functionalists, by extending the Homunculus head analogy. It seems possible that you could imagine a large enough number of people performing the tasks of those little men, all connected up by radio-link with one-another, and all performing operations according to bulletin boards held in space upon satellites. For a short while, it ought to be possible to replicate a piece of person's behaviour in virtue of making the connections between the inputs and outputs according to a machine table. If this population of people is capable of instantiating a mind, then there should be something akin to felt experience by this population. Block's point is that it is counter-intuitive to say the least, that one would know how to understand this suggestion.

### Subsection 2 : Are inverted qualia possible?

A more far-reaching objection involves the possibility of qualia inversion. Consider two people presented with a green object, such as a field of spring barley on an April morning. Both would describe their experience as being of a patch of green. This mental state would serve to explain their behaviour with respect to the field; they would both say that it was the same colour as the 'Go' signal on traffic lights, that it falls within a range of wave lengths, and that the colour had certain psychological properties, such as making them feel calm. To the extent that the experiences would be functionally identical, so too would the mental state types be identical between the individuals. However, if it is possible that the experience one individual has in such a situation, is the same experience as the other has when seeing something red, then the mental state types are distinct. Block argues that, in defining mental states as processes that associate inputs to outputs, Functionalism is incapable of accounting for what is distinctive about the nature of some of our mental states - in particular, the experiences linked to our sense organs. As there is no functional difference between the two colours on this account, inverted qualia seems to be an instance of a mental state that falls outside the Functional analysis.

### Subsection 3 : Implications of the above possibilities for Functionalism.

The moral of the above considerations is that if we use Functionalism to arbitrate upon claims of the existence of mental states, then there will be counter-intuitive results. Block introduces the terms 'Liberalism' and 'Chauvinism' to indicate defects in criteria for differentiating between objects with minds, and those without.

Where any theory, Behaviourism for instance, allows for the attribution of mental states to objects which would otherwise be straightforwardly considered inert, then that theory may be described as 'Liberal'. On the other hand, where any theory, such as the Identity Theory, withholds attribution of mental states to objects which we would normally think of as sentient, then that theory would be described as 'Chauvinistic'. He concludes that the Functionalists will be unable to specify the inputs and outputs to any process, in such a way that steers a line between liberalism and chauvinism. What is required, is some restraint upon what would count as a suitable input or output, which is not so abstract as to include such entities as 'the economy of Bolivia', but which is not so specific as to preclude other possible species. Yet, can this be done in a Materialist framework? Block suggests not:

The functionalist says to the physicalist: "It is very hard to see how there could be a single physical characterization of the internal states of all and only creatures with mentality." I say to the functionalist: "It is very hard to see how there could be a single physical characterization of the inputs and outputs of all and only creatures with mentality."<sup>12</sup>

### Section Three

However, many have adjudged Functionalism too good a theory to abandon without a fight. Elliott Sober and William Lycan particularly, have taken up Block's challenge, and have refined Functionalism in a manner that was expected to settle many of Block's qualms.

#### Subsection 1 : Sober - Functions as 'purposes'.

Sober notes that the term 'function' has two distinct uses. As it appears in mathematics, functions specify how one object should be transformed by another. Yet it also has a teleological meaning, which describes the purpose of the object. Because of the persuasiveness of the computer model of functionality, the mathematical meaning of 'function' had completely overshadowed the alternative construal. However, once the term is used to denote 'purpose', the above objections become less pressing.

'Function' is a rather vague term used on its own. To do any work, there has to be a specification of the relation a feature, or process, has to the object for which it is a function: that is, everything can be said to have a function - the ascription of the term only becomes explanatory once it is relativized.

Functionalists hope to characterize psychological states in terms of their causal connections to behaviour, to stimuli, and to other mental states. Only

---

<sup>12</sup> Block. 1978. P 465.

insofar as these provide substantive constraints on how psychological states are individuated does functionalism become a non-trivial doctrine.<sup>13</sup>

To illustrate how we might constrain the application of mental states, Sober draws an analogy between mental processes and digestive processes. Digestion is characterised in terms of its purpose. In its most abstract sense, digestion provides the organism with the resources necessary to expend energy, maintain and replace tissue. From this very general description, it is clear that mechanisms of digestion will vary from one organism to another, depending upon their nutritional requirements and their physical composition. In this sense, it is unreasonable to expect all and only those digestive systems like ours to count as organisms having digestive systems. Similarly with the mind. Mental states such as pain have a purpose, which is to protect the organism from harming itself fatally. So long as there is some discernible mechanism in the organism which serves this purpose, then that mechanism need not resemble ours at all. Block has made the mistake of assuming that for each input, there is a determinate output; an assumption that is sanctioned by the mathematical use of the term, but not so with the teleological usage.

### Subsection 2 : Lycan - 'Homuncular Functionalism'.

Lycan similarly identifies the problems as stemming from concentrating too hard on a computer-style form of Functionalism. He acknowledges that such a metaphor neatly shows the distinction between rôle and occupant, where any particular task - or rôle - can be served by a number of programs - occupiers. However, there is a danger that this may lead to the assumption that nature conforms to this two-layer pattern. He does not want to deny this distinction exists; what he objects to, is that it should be seen as a binary 'either-or' distinction. Instead, he suggests that the relation should be thought of as lying along a continuum. For instance, consider a cell. It may be thought of as occupying a rôle relative to any piece of tissue or organ of which it is a part. Alternatively, the contents of the cell may be thought of as occupying a rôle relative to the cell of which they are a part. Generally the rôle will be specified in fairly abstract terms, such as 'to serve as a barrier', with a more specific physical description of how the occupant fulfils this rôle. Importantly, this distinction does not imply that objects naturally fall on one or other side of the divide. Rather, they are distinguished according to how they are related along a continuum of increasing complexity, running from the specification of basic physical properties, such as 'having a negative charge', to highly abstract properties, such as 'regretting an action'.

Daniel Dennett, suggests a way in which purposeful activities are analysed in Artificial Intelligence, can provide a helpful model of how

---

<sup>13</sup> Sober. 1985. P 176.

mental processes are performed by the brain<sup>14</sup>. Typically, when designing programs that attempt to simulate cognitive processes, behaviour is broken down to lower and lower levels of abstraction, until a level of specific action is reached. Tasks are described by a flowcharts, with black boxes representing sub-processes. These boxes are themselves represented in a flowchart, with further use of black-boxes, which are in turn represented by further flowcharts. The idea is that each layer of representation constitutes a layer of less abstraction and greater mechanical specificity; what starts off as an intelligent process, is broken down into progressively simple operations, such as comparing two memory registers.

For instance, consider how this technique might be applied to food spotting tasks. At a highly abstract level, one describes first the discrimination of objects in the environment. This process can be subdivided into perceptual and recognition tasks; these into signalling functions; and further down, into ways cells transmit electrical impulses. The Materialist wager is that there will be a seamless transition from the highly abstract vocabulary of mental states, to a more specific description of how organic matter reacts in certain ways in certain circumstances.

Lycan calls this 'Homuncular Functionalism', adverting to the idea of sub-systems performing tasks of increasing simplicity, down to the level of basic chemistry. As with Sober's view, this picture is faithful to the idea of multi-realizability. The highly abstract level of feeling pain, for instance, can be analysed according to a number of sub-routines. This reflects the idea that organisms physiologically quite different from ourselves, can nonetheless have pain states. This revision of Functionalism results in the following analysis of mental states:

I propose to type-identify a mental state with the property of having such-and-such an institutionally characterized state of affairs obtaining in one (or more) of one's appropriate homunctional departments or subagencies.<sup>15</sup>

There are a number of advantages to this version of Functionalism. The hierarchical structure of sub-processes eludes a common failing, that was pointed out by Ryle<sup>16</sup>. In 'The Concept of Mind', Ryle showed that with cognition and sensation, any theory that attempted to explain one process as being part of a prior process, would involve itself in problems of infinite regress; thus an act of deduction could not be explained by an inner act of thinking, nor could an inner eye scrutinise our perceptions. However, Lycan claims that because the sub-systems are a refinement rather than a re-definition, the threat of regress disappears. The sub-systems perform a different

---

<sup>14</sup> Dennett. 1975.

<sup>15</sup> Lycan. 1987. P 41.

<sup>16</sup> Ryle. 1949. Chapter Two, Pp 21-22. Chapter Six, P 156.





task - typically, one which is less intelligent - so their existence in a theory is genuinely doing some explanatory work, rather than merely pushing the problem further into the background, in the hope that it will eventually disappear.

The idea that mental processes can be broken down into components appears to have empirical support. As an example, Lycan cites the time-lag between the realisation that you have been burnt, and an intense sensation of pain. Pain appears to be divisible into a sensation and typical behaviour. This apparent division between various aspects of the processing of bodily harm, is supported by an experiment performed on subjects being treated with different drugs. With one drug, the patients report that the pain had disappeared completely; with a different drug, the pain was said to be there, though patients could not 'feel' it; with a third drug, the pain was as intense as ever, but patients did not find it bothersome at all. A plausible interpretation of these results suggests that different processes normally combine together to give typical pain behaviour, but that as the constituent parts are inhibited with drugs, abnormal and seemingly unintelligible results emerge - such as it being possible to experience intense pain and not be concerned.

If this analysis is correct, then it provides an explanation for why it is that the nature of mental states seem to be so far removed from the physical. The difference is greatest at the extremes of the continuum, between the abstract and the specific. However, as one progresses along the line, there are points at which it is hard to tell whether the processing is best described as metaphorical or mechanical. By thinking in terms of two levels of nature, this picture is encouraged. Instead, perhaps, we should think that the grading of mental states into physical states, as being similar to that of the grading of one colour into another; at each end of the spectrum the colours bear no resemblance, whereas at adjacent points, it is difficult to say whether the hues are definitely of different colours.

### Subsection 3 : 'Function' as a Teleological concept.

Functionalists such as Sober and Lycan, realised that describing mental states as functional, added nothing to the debate: there is a sense in which everything plays some sort of function, and in some ways, both Behaviourism and the Identity Theory are functional theories. What was required, was some fleshing-out of the particular way that the term 'function' was being used that would sanction Materialism. To that end, Functionalism has looked toward theories of natural selection. The function of biological types is specified in terms of how that type contributes to the survival of the organism in which it appears. Thus, the explanation of the origin of the opposable thumb is couched in terms of the causal abilities the feature confers. 'Function' has come to be identified with the consequences of the presence or

absence of a particular feature, and this has become the standard way of interpreting function as applied to mental states.

‘Aetiological’ functionalism was originally articulated by Larry Wright, in an article titled “Functions”. There he argued for a particular analysis of how properly to attribute functions to specific features of an organism:

When we explain the presence or existence of *X* by appeal to a consequence *Z*, the overriding consideration is that *Z* must be or create conditions conducive to the survival or maintenance of *X*.<sup>17</sup>

However, there are some doubts concerning the soundness of this analysis. The problem is that aetiological explanation is essentially historical. It explains the existence of any particular feature of an organism according to how that feature contributes to the fitness of members of that species over time. Clearly, this restricts the range of possible explanations to the past. The theory requires some amendment if it is to have any application to the current and future of species. Both in ‘Fitness and Function’, and ‘Wide Content Individualism’, Denis Walsh points to two major difficulties to which this restriction gives rise. Firstly, by being tied to the past record of success, it is unexplained how novel features can still be relevant to survival, let alone how they continue to be relevant. Not only do we explain how new features appear in terms of how they benefit the species, but also we are able to predict how species might change with their environments. The other defect in the account is that it fails to provide criteria for normativity with respect to particular features. As Walsh observes, my liver ought to function in a particular way, because of the effect that it has on my existence, not because that is the way it contributed to my ancestors’ survival. Nor should I acknowledge the importance of my appendix, simply because in the past it served an essential purpose.

Ideally, the term ‘function’ has to relate as much to the present individual as to the past succession of species. Walsh’s solution is recognise that fitness is a condition that is relative to the environment in which the organism finds itself, and proposes to amend the aetiological theory with a ‘Relational’ theory:

The leading idea of the relational theory, like the aetiological theory, is that the function of a trait is tied to the way traits of its type contribute to the individual fitness. However, the way a trait contributes to fitness varies from context to context.<sup>18</sup>

With this amendment in place, the problems cited above dissolve. With any purported explanation of a function, it has to be specified

---

<sup>17</sup> Wright. 1973. P 164.

<sup>18</sup> Walsh. 1998. P 633.

relative to the particular context - whether the past, present or the future is being considered. This is only to be expected given that as circumstances change, so too will the relevance of a particular trait. Accordingly, for any trait token, its function will have to be related to the context in which it is found, in order to determine its importance. For instance, the presence of certain reflexes in humans are explained in terms of early evolutionary needs, that are no longer relevant. Their being present in humans is not explained by their contributing to fitness, since human survival no longer depends upon those particular traits. Their presence is explained by their being relevant once, but as the circumstances have changed, so too has their significance. These are features that we could reasonably expect to evolve out of the system, as they no longer provide an advantage to their possessors over other members of the species.

The Relational theory also satisfies the normative element. By making fitness context relative, it is possible to compare past and present contexts to determine whether there is a reason why my liver or appendix may be considered useful. Further, the theory offers criteria for discrimination between features which are genuinely useful to a species, from those which are accidentally useful. This is done by averaging-out the contribution of a trait's usefulness to a range of individuals within a particular context. Where an individual has a particular trait which does not contribute to the average fitness of the species, then it may be regarded as a freak occurrence.

#### Subsection 4 : Functionalist solutions to the qualia problems.

Pro-Functionalists believe that they are now in a position to answer Block's objections. Once 'purpose' is employed within the analysis of the relation between mental states and physical states, then the machine-table counter-examples miss the point. The interaction between the states has to have some way in which it serves the interests of the organism for which it is functioning. A Functionalist now differentiates between two notions of function, and defines the purpose of mental states to contribute to an organism's biological needs. Thus, cases of analogous processes such as 'the Chinese nation', fail to embarrass Functionalism, because they cannot be regarded as processes serving any particular needs.

The extent to which we can accommodate artificial devices within our concept of sentience, is not something that is clear-cut. Sober illustrates this point with the example of a pacemaker attached to the heart. We would not deny that the heart fulfils its purpose even though it is supplemented with the pacemaker. However, if we were to attach the pacemaker to a computer, then although it would still be serving as a regulator, there is no reason for thinking that is performing the same function as with the heart. Similarly, were a program to perform the task of parts of the brain, we might want to

attribute some mental states to the person involved. This does not mean that we would want to attribute those states to any machine upon which those programs were run. Just as a machine does not become a 'body', simply because it has a pacemaker attached; neither does a machine become a 'brain' in virtue of its running such a program.

So far as the inverted qualia problem is concerned, Sober points out that it is one thing to be functionally equivalent, but another to be biologically identical. Certainly, Functionalism is unable to point to any functional difference between two individuals facing a patch of blue, even though their experiences are quite distinct. He doubts whether they will be physiologically similar though. Whilst there is no functional difference to a rose whether its thorns are large and regularly spaced, or small, thin and multitudinous, there will be a biological difference to distinguish one rose from another. Block's mistake is to infer from the fact of functional similarity, that there will be biological similarity.

Against the problem of input and output that Block raised, Lycan attempts to defend Functionalism with the three following points. Firstly, he notes that Block assumes that all inputs and outputs have to be described at the same level of abstraction. However, if there is such a continuum as discussed above, then there is a possibility that some descriptions of suitable inputs and outputs might be abstract enough to avoid the charge of chauvinism and yet be strict enough to be able to rule out Bolivian economies. Secondly, some mental states appear to be more abstract than others, such as 'fear', whereas others seem to be more specific, such as qualitative states generally. One might expect all creatures with minds to have some process which produced typical fear behaviour, and thereby allow for a diversity of realization of this process. Alternatively, it is far from clear that we should expect colour discrimination to be a necessary part of all sentient creatures, and so need not look for mechanisms which could perform such processing. Thirdly, there might be a case for allowing that the nation does have certain dispositional mental states such as beliefs, as well as stating that there are some qualitative states which will only be experienced by organisms with such an anatomy.

In conclusion, the Functionalist position has evolved from a theory that relates inputs and outputs by the mediation of mental states in the manner of some machines, to a theory that sees mental states as being traits of an organism which have been advantageous in the past, and which continue to confer advantage. On this basis, mental states are simply evolved behavioural traits that enable the species to survive in the present environment.

#### Section Four

It should come as no surprise to learn that such a conclusion is not so much the culmination of a philosophical enquiry, but more of a

spring-board for further discussion. As yet, there is an inadequate understanding of the rôle played by the *character* of sensations. Whilst one should accept that the teleological aspect of the position has blunted the force of some of Block's absent qualia objections, there is still some doubt about whether or not the Functionalists have appreciated the force of the inverted qualia examples. In fact, the status of qualia seems to have received a raw deal with Functionalism. This Section considers two instances that support the above accusation.

### Subsection 1 : Can Functionalism deal with 'inverted qualia'?

In order to establish that inverted qualia are an impossibility, Functionalists adduce evidence from psychology - for example, the extra-chromatic effects red and green have. There are many psychological corollaries with wave-length such as make it implausible that you could have a different reaction to that particular wave-length, and hence that particular colour. In this sense, it is suggested that colour is important to us, not merely as being a property of objects, but also as playing a rôle in our concepts of emotion. For instance, there are cultural associations between blue and despair, white and purity, red with anger and black with mourning. On this account, if the colour sensation experienced were inverted from red to green or blue to yellow, then an individual would then find it hard to see the sense of describing someone as 'feeling blue', if that meant experiencing a rather upbeat shade of yellow. There are further associations which have become cultural norms in the other modalities; for example, the touch of cold porridge is always unpleasantly slimy, the sound of the babbling brook is proverbially soporific, the taste of lemon could only be sharp, and the smell of decay rather nauseous. The possibility of inverted qualia, is refuted by a 'consistency thesis'. If there were the degree of variation between people's colour sensations consonant with the inverted qualia hypothesis, then it would be impossible to develop cultural associations such as between colours and moods, for instance. The presence of these affiliations, proves the consistency of people's qualitative experiences.

Some of our associations have been explained in terms of the physical properties of our senses. The 'Lightness' of yellows and reds and the 'Darkness' of blues and greens, can be explained in terms of the eyes sensitivity to different wave-lengths, with those bands of the spectrum towards the red end being more sensitive than the blue end. Similarly, the appearance of some colours as receding or advancing, can be explained in terms of the lens focusing light of different wave-lengths at different points relative to the retina.

These reassurances miss the point. It is one thing to associate wave-length with emotional effect in these terms, but what remains unanswered is how that wave-length corresponds with the quality of the sensation as experienced by me. What is left unproved, is whether the

qualia influence the emotions or whether the wave-length of light has a direct effect upon the emotions. Further, it may be plausible for colours that are opposites, but the above argument against inversion is less plausible when considering two colours that are closer together on the spectrum - such as the green of verdigris and the blue of a Dunnock's egg. It is insufficient to argue upon the basis of psychological characteristics here, since there are no gross psychological correlations between colours as finely shaded as these; yet, the blue and the green are qualitatively different. One might concede that our association of terms would not allow a dramatic variation in qualia, such as one person experiencing yellow where another experienced blue: but it does not follow that we have to concede the point for all colour differences whatever.

Associations are in some sense an expression of a value, rather than a feature of the object. For example, cultural associations with a colour in one context, are incompatible with the same colour in a different context. Thus red is both related to warmth and security, as well as with danger and anger. Blue with melancholia and calmness; white both with purity and ghosts. If there is this lack of consistency amongst the association of colour terms, then the Functionalist will have to show how it is compatible with the consistency thesis, introduced to anchor colour qualia to just one uniform type. The consistency thesis holds that the possibility of there being these cultural connections between colours and behaviour, follows from there being no instances of inverted qualia in that community; inversions would show up against this cultural background. This ignores the fact that there is inconsistency with respect to the affect colours have upon us, so the thesis fails to guarantee that there cannot be inverted qualia. If there is no consistency amongst the emotions aroused by a single colour, it does suggest that there is more to the physical aspects of a colour sensation than is given on the Functionalist account.

### Subsection 2 : Does Functionalism explain why we have sensations?

What is the function of 'pain'? Lycan addresses this question in a footnote<sup>19</sup>; to him it is clear why humans should have a such a sensation. The argument runs as follows. It is the function of pain to alert the organism to the possibility of damage to its body. Thereby, the organism can prevent what is damaging it, and take steps to repair that damage. The question is then re-phrased - why do we need this to be done by 'pain'? Why not utilise a set of propositions, made available to introspection, alerting the individual to the fact that it would be well advised to perform a damage avoidance routine? That is to say, in the relation between sensory input and beliefs about the

---

<sup>19</sup> Lycan. 1987. P 138, footnote 17.

environment, why does information have to be conveyed by a sensation of some sort?

The suggested answer is that 'pain' is a state that provides a stronger motive to act, in comparison with a thought. This reply is backed-up with an analogy between an experience of a sharp pain and a warning system - such as a light that comes on when batteries are low on a computer. The red light gives adequate warning of something that is going to occur, and provides us with a good reason to perform certain actions - such as changing the batteries. However, the only problem with this sort of system, is that it is too easy to ignore the warning, and find oneself in trouble at a later date. With something like pain, there is no possibility of overlooking it: that is why fire alarms are shrill and difficult to ignore - a system which consisted of a single bulb that illuminated in times of crisis, would hardly have the dramatic import that the possibility of being burnt alive should have.

There are two reasons for gazing askance at this argument. One is that, even were it conceded as being conclusive for explaining why *pains* feel like they do, it still has not given an adequate response to why the other sensation types have their particular modes. Take for instance something lightly scented, such as the odour of copper coins. This sensation is as easily ignored as the warning light was supposed to be. There is no advantage to be gained by having a sensation of this property, other than the thought that there is copper in the vicinity. The idea that we have sensations rather than propositions because they are more emphatic, and harder to ignore, is not supported by the above instance, and leaves the particular properties of at least some sensations without any clear function; their task would be as well served by any other sensation type, or by no sensation at all.

Secondly, why accept this account even of pain? Not all inimical faults are painful. Not all painful situations are dangerous; often one is astonished at how intense is the agony for so minor an ailment, such as the discomfort caused by an accumulation of wind. Further, like the red warning light, it is quite easy to ignore some pains as well, admittedly incurring the problem of aggravated injury later on. On the face of it, the Functionalist has yet to account for why the felt qualities of sensations, should be so important to human existence.

### Section Five

Fortunately for the Functionalist, there is a resource at hand; a theory of perceptual content called variously Representationalism, or Intentionalism. These theories are a natural extension of Functionalism. Representational theories have been used to explain the nature of all Mental states; not just sensations, but also those which arguably have no character whatsoever, such as thoughts. However, for the purposes of this chapter, it is sufficient to consider solely those aspects of Representationalism that have to do with sensations.



### Subsection 1 : The appeal of Representationalism.

Such theories appear in their most attractive guise when giving an account of sensory experience, since such theories explain the following seemingly odd situation. It has been an unhappy consequence of the discovery of the atom, that the following is bound to occur during the early part of one's childhood. Someone will tell you that what one takes to be a perfectly solid looking chair, is in fact nothing more than a whirling, gappy, mass of particles; that the surface of the desk is far from smooth, but is a rugged and discontinuous terrain; that the sky is not really blue; in fact, anything that hitherto made life worthwhile, is just an illusion and our senses are not to be trusted. However, one can find solace in the idea that our senses represent the world as having colour and shape, as being fragrant, harmonious, smooth and tasty, without our having to worry whether the world really is like that. According to Frank Jackson<sup>20</sup>, the attraction of Representationalism, is that it bridges the gap between the nature of the world as appreciated by our senses, with the nature of the world as described by our sciences.

When we perceive, it is not the object that we perceive directly, but a representation of that object. Accordingly, all our sensations are representations, providing information about states of affairs either in the outside world, if they are via the senses, or states of affairs internally, if they are bodily sensations or emotions. Michael Tye articulates the Representationalist position as claiming:

...that experiences and feelings are sensory representations that elicit various sorts of cognitive reactions, and that differences in what the sensory representations represent, go along with differences in what it is like to undergo the experiences and feelings.<sup>21</sup>

With the idea of computers permanently swilling around in the background, it is not surprising that Functionalists should find the idea that feelings are representations congenial, or that Representationalists should resort to descriptions that have a functional slant. Much work has been done to articulate an analysis of 'thinking' in terms of symbol manipulation. A computer system is regarded as a paradigm information processor, so it is natural to think of inputs as having a representational form. In some sense, inputs and outputs have to be compatible with the sort of processing that any system performs. Programs are designed to anticipate a certain form of input, and will attempt to interpret what they have to deal with in terms of these in-built expectancies. It is the parameters of the processing software that determine whether or not input is suitable. According to how successful it is in its interpretation, the system will either process the

---

<sup>20</sup> Jackson. 1993. P 187.

<sup>21</sup> Tye. 1995. P 68.



data satisfactorily, produce complete nonsense, or reject it outright.

The alluring advantage of Representationalist views is that they present sensory data as, and only as, information; there is nothing outside the experience that is not a representation of the external object - in the case of sight, sound, taste, smell and touch - or an internal one - in the case of bodily feelings. This characterisation of experience symbolically, as it were, is perfectly aligned to be considered as providing the 'input' into the Functionalist's account of the Mental.

### Subsection 2 : The relation between Functionalism and Representationalism.

However, it is one thing to appreciate how nicely the theories compliment one another, but it is another to demonstrate that some form of Representationalism is forced upon the Functionalist. There are at least three objections that suggest the two positions are independent. Arguably, none of them are successful. It might be objected that so far, the senses are being regarded as a form of input, whereas it might be more accurate to say that a lot of processing goes on before one has the experience of, for example, seeing a sea-washed pebble as 'smooth'. Why not regard these experiences as a form of output? The Functionalist is looking for an explanation of how experiences interact with other mental states, and is quite happy for the sensation of a smooth pebble to be an output of a process, since such an output can stand as further input. It is argued that, after such processing, the initial input is transformed into the appropriate state to serve as input to a further process, such as the forming of a belief that the pebble has lain on the beach for many years.

A second objection questions the assumption that all inputs must be representational. There are many other types of processing, whose inputs do not appear to have an interpreted nature. The existence of such processes suggests that it is not simply in virtue of being part of a process, that an input for that process is thereby a representation of an object for that process. If it were, then one would be committed to ascribing mental states to all sorts of devices, such as toasters - after all, they process the bread to degrees of burning and the inputs are bread and heat, without there being any sane suggestion that these inputs are represented to the toaster in some recondite manner. The reply to this objection refers to the Teleological element of Functionalism, whereby it is not just any old system with which they are concerned, but only those with some recognisably biological purposes.

There is one other potential explanation for the nature of the character of sensations, which does not construe it in terms of representations. Such character is a manifestation of the properties of a Mental substance, that is completely independent of the physical

nature of things. The fragrance of a particular rose is a result of the chemical properties of the rose interacting with my mind. In this case, the particular fragrance is not a representation of anything; it is simply the scent of a rose. Clearly, a Functionalist would not wish to embrace this form of explanation, however dire the consequences of accepting a Representationalist theory, since to do so, would be to accept a Dualistic position - the very position that Functionalism was devised to replace.

It looks therefore as if the Functionalist is left with no choice but to construe phenomenal character in terms of Representations of sensible properties of the world. To many philosophers, Representationalism is regarded as the most promising form of Materialism to date, so the above conclusion will not cause them any anxiety.

### Section Six

Many believe that Materialism's only chance of giving a convincing account of the mind lies in embracing some form of Representationalism. There are, however, some reasons for thinking that this belief is mistaken. This section deals briefly with the idea that the motivation behind such a theory is not as strong as philosophers assume.

In the article "What Is a Theory of Mental Representation?", Steven Stich argues for this position. He notes that:

The quest for an adequate theory of mental representation is not just a popular pursuit, many writers insist, it is a vitally important one.<sup>22</sup>

There are two connected strands in Stich's article; that there is no obviously 'philosophical' rôle in our attempt to understand how we attribute mental states; that the threat of Eliminativism is far less forceful than we are led to believe. Instead of feverishly shoring-up deficiencies in Representationalism, philosophers should relax, and let a discipline such as Cognitive science investigate the mechanisms by which we come to 'entertain beliefs', 'succumb to fears' or 'perceive that something is salty'.

He suggests that the desire for a theory of mental representation, arises from a fear of Eliminativism. Unless we can justify our use of folk psychological terms, there is a danger that these terms will be removed from our discourse, just in the way that we no longer talk about the Fates spinning out our destiny. One way of providing this justification, is by finding scientifically respectable correlates for these folk terms. In this manner, our common-sense use of mental vocabulary would mature into a rigorous scientific theory - immune from the brutish tendencies of Eliminativists. As a candidate for this

---

<sup>22</sup> Stich. 1992. P 243.

theory, Representationalism seems ideally placed. However, as will become evident below, there are many distinct ways of articulating a theory of mental representation: so, how should the correct theory be chosen?

Subsection 1 : Is philosophy the correct discipline for understanding the mind?

It is at this point that philosophic interest is engaged. In order to integrate the two versions of our mental vocabulary, we require a clear idea of the nature of the concepts involved. One way of doing this, is to employ the Socratic method of finding necessary and sufficient conditions for the application of the concept in question; there are two reasons for thinking that this particular method is unsuitable. Firstly, when this method has been used in the past, it has met with very little success. This is not surprising when one considers the following. The success of this method is at the mercy of the competence of the individuals engaged in the analysis. Ideally they will know whether any proposed formulation admits of a counter-example. This is to admit that the method will work only in those cases where the individuals have prior knowledge of the necessary and sufficient conditions that they are attempting to elucidate; which is to say, either the method will never work to anyone's knowledge - since they will not be wise enough to realise that they have actually found the correct conditions - or that the method is redundant - since such conditions are knowable independently of the method being used. That is not a reason for thinking that such a method would not be fruitful on this occasion, but prudence dictates that some alternative be considered.

Secondly, research into the patterns of concept use, show that people neither create nor employ their concepts in a way that is compatible with the property of having necessary and sufficient conditions of application. Instead, they use a hodgepodge of devices, such as stereotypes, and spur of the moment judgements. The fact that such concepts are not amenable to the Socratic method, does not entail that they cannot be involved in respectable theories. For instance, the nature of linguistic practices has increased in spite of its employing methods that eschewed the notions of necessary and sufficient conditions. Comparing the relative successes that these alternative methods have enjoyed, it would appear that the philosophic approach should give way to the Cognitive Sciences. Stich concludes that, if the Socratic method is the only contribution philosophy has to make, then it has no rôle to play in the devising of a theory of mental representation.

Subsection 2 : Are there reasonable grounds for fearing  
Eliminativism?

This leaves the threat of elimination of at least some of the terms of our folk psychology. If we devise a successful scientific theory, then we can discard any expressions from every-day discourse that do not figure in that science. However, one reason that Materialism has such a hard row to hoe, is because mental terms are so difficult to merge with those terms of the physical sciences. Whatever else is left after folk psychology is purged of scientifically unacceptable items, it will not include such things as are recognisably mental states from the first person perspective. So perhaps it is worthwhile to try again and acquire a theory that does respect the first person sense of mentality.

Stich doubts the danger is so great as to warrant the devising of so improbable a theory. The threat of Eliminativism is at its strongest when there is only one theory that is going to replace folk psychology. Yet, there are many such contenders, and they are not all of them mutually exclusive: that is, the truth of one does not entail the falsity of another. They are not all in competition, and need not be using terms univocally. Therefore, in order for elimination to be a strong possibility, there will have to be a choice of which is to be the theory that is to replace our every-day use of mental terms. Even were it possible to make this choice, there appears to be a grave obstacle in the shape of determining the terms which are ripe for elimination.

Most Eliminativist's have chosen a descriptive theory of reference; if two descriptions match, then they refer to the same object. However, there is a problem with knowing how to determine the conditions under which one is justified in saying two descriptions match. If the criteria are too strict, then a consequence would be the impossibility of ever increasing our understanding. For example, a team of scientists are investigating the physiology of swans. They learn that it is possible for black swans to exist. Clearly, their use of the term 'swan' is subtly different once they have acquired this new information. Whilst common-sense suggests that such a difference does not entail that they are now studying a completely different creature, it is unclear at what point can one allow new information to have no effect on the terms used. For at the other extreme, one cannot be too lax without making all theories equivalent; there has to be some justification for denying that the phlogiston theory of oxidation refers to the same processes and objects as the present day theory. It is essential that some point be chosen along this line from strict to lax, if there is to be the elimination programme envisaged.

Eliminativism will be associated with some form of theory of reference. However, there will be competing theories for this rôle, so the difficulty of having to choose between theories returns at a different level. Until a theory of reference is decided upon, it is impossible to say how much of our common-sense mental vocabulary is under threat.

Granted the problems involved in the selection of an appropriate theory, it is a little premature to assume that even were Eliminativism to get off the ground, it is bound to be iconoclastic. The threat of elimination is not that great.

### Subsection 3 : Can we live without Naturalising the mind?

We seem to be afraid of the consequences of finding that there cannot be a naturalisation of intentionality. Stich believes we should ask ourselves two things: firstly, how we anticipate a theory of mental representation will address this fear; secondly, just how unbearable will the failure of this project be? He anticipates that our response to the first question will be that a theory of mental representation will embody a means by which we can preserve our common-sense beliefs in the mind, by embedding them in a suitable scientific discipline. Yet faith in the success of this project is seriously undermined by the points made earlier about the identification of theoretic concepts, and the struggle to find necessary and sufficient conditions. Which leads to a response to the second question; if the project is in jeopardy, so what? It is simply false to conclude that any term lacking in Naturalistic credentials is thereby devoid of scientific utility. He points to a number of cases that establish this point - for instance, the notions of 'grooming' and 'attack behaviour' as used in the studies of primates. These notions are employed in a respectable scientific discipline. They lack the status of naturalised concepts. Would gaining this status make any relevant difference? Stich thinks not:

To demand more - in particular to demand that the notions in question can be "naturalized" - seems unmotivated and silly. The situation for *mental representation* looks entirely parallel.<sup>23</sup>

There is, therefore, no real need for a theory of mental representation. If it is thought desirable because it promises to provide necessary and sufficient conditions for our mental concepts, then there is every expectation that this promise will remain unfulfilled. If it is wanted to ward off the possible elimination of these concepts, then at least let us wait until the Eliminativist programme has got itself in order, before judging the severity of this threat. Why assume it is needed to lend scientific credibility to mental terms, when its absence has not prevented other notions enjoying a useful rôle in science? What other reason could there be? The struggle to find an appropriate theory is just so much misplaced energy. Our understanding of the mind is better served by the disciplines of Cognitive Science.

---

<sup>23</sup> Stich. 1992. P 258.

#### Subsection 4 : Response to Stich.

Stich has campaigned for the removal of mental states from figuring in the work of scientific understanding of behaviour. It should come as no surprise that he is not overwhelmed with excitement at the prospect of another theory that claims to deal with just that subject. Regardless of the cogency of his arguments above, it is still possible to retain some interest in a Representational theory of mind. Although he enquires: "why do we want one?", he ignores the obvious reason that, without such a theory, there is no reason for thinking that the mind and the body are importantly connected. We do not want to be told that there are various scientific disciplines that will provide explanations of behaviour patterns. What we want to know is how the physical matter of the central nervous system is capable of realizing mental states. Without this theory, we can not progress to understand how it is that mind and body interact, nor the extent to which other organisms are capable of sensations or thoughts. Any breakthrough that occurs to explain how the mind and the body are related, will supplement the explanations furnished by Cognitive Science. It cannot be satisfactory to suppose that a discipline that is indifferent to how the mind and body are connected, will provide answers to the questions that this thesis set itself to explore. In which case, whether or not it is a theory of Representationalism that provides us with the answers, it is still important that philosophic enquiry continue, whilst there is still doubt as to the nature of the mind-body relationship.

#### Section Seven

It appears that if we want a theory to explain how it is that the mental and the physical are one, then we had better pursue Representationalism. This Section develops the theory before criticising a fully fledged version in the next chapter. Once the position is defined, it emerges that there are two discernible levels of consciousness. The reasoning behind this distinction is given in Subsection 1. With this distinction, comes a division upon how to give an account of the nature of consciousness; whether it is best understood through a 'Vertical' theory, or whether a 'Horizontal' theory is superior. In the remainder of the Section, the two 'Vertical' theories will be discussed and then rejected. The 'Horizontal' theory, will be the subject of Chapter Four below.

In "Consciousness and Experience", William Lycan makes the following claim:

...the mind has no special properties that are not exhausted by its representational properties, along with or in combination with the functional organization of its components. It would follow that once representation itself is (eventually) understood, then not only consciousness...but subjectivity,

qualia, “what it’s like”, and every other aspect of the mental will be explicable in terms of representation together with the underlying functionally organized neuro-physiology, without our positing any other ingredient not already understood from the naturalistic point of view.<sup>24</sup>

One reason why Materialists might be excused getting excited about this idea, is because it promises to provide a naturalistic account of ‘intentionality’. As noted in Chapter One, the property of being able to think about and imagine, states of affairs that do not exist, is a feature of the mind that resists reduction to the material. In any other relation between physical objects, such as being ‘beside’, ‘against’, ‘taller than’ and so forth, the objects specified in the relationship must exist if the ascription is to make any sense. Yet, it is possible for the mind to represent non-existent objects, such as when one believes that the Thirtieth of February is a Tuesday, or when one hallucinates a distant oasis whilst wandering in a desert. The fact that the mind is capable of this sort of relation, and that no other physical entity is, provides a difficulty for anyone trying to accommodate the mind within a Materialist framework. This difficulty has become known as ‘Brentano’s Problem’<sup>25</sup>, after the Nineteenth century philosopher Franz Brentano, who argued that ‘intentionality’ was the defining characteristic of mental phenomena.<sup>26</sup>

The way Representationalism addresses this challenge, is to hold that the mind may usefully be divided into a sensitive side and a cognitive side. The sensations are the product of stimuli, and their character is systematically associated with the natural properties to which organisms are sensitive. In this sense, the character of a particular sensation represents the existence of some property, either in the world external to our bodies, in the case of the five senses, or of states of affairs within our bodies, in the case of pains, emotions, and moods. The qualities of these sensations are used by the cognitive side of the mind, to form beliefs about the creature’s environment, in order to satisfy such desires as for food, shelter and freedom from predators. In this manner, the character of experience is regarded as a natural relation between the physiological properties of the sense organs, and the properties of objects to which they are sensitive; and the nature of cognitive processes, with which there is arguably no associated ‘feel’, is regarded as being analogous to those processes of symbolic code manipulation as are carried out by computers.

One might well wonder why it is important to discuss three Representational theories separately, instead of simply considering what they have in common. However, it is possible to make a distinction between them, and the principle by which they may be

---

<sup>24</sup> Lycan. 1996. P 11.

<sup>25</sup> See Tye 1994.

<sup>26</sup> See Brentano 1873. Book 2, Chapter 1.

differentiated has implications for the success of any such theory. The basis for this distinction is described by Fred Dretske, particularly in the article "Conscious Experience"<sup>27</sup> and the book "Naturalizing the Mind"<sup>28</sup>; the following discussion is drawn from these sources.

Subsection 1 : The notion of two levels of consciousness.

What separates these theories, is the way in which they account for the relation between representations and our awareness of them. Intuitively, it appears obvious that we can experience sights and sounds without our being aware that we are doing so. This type of occurrence is clearly illustrated by David Armstrong, in the essay "What is Consciousness?" He considers the case of a long-distance lorry driver, who suddenly realises that he has been driving for some time, but is unable to recollect any details of his journey over that period. What are we to say about the driver's state of consciousness during that time? Armstrong suggests the following:

There was mental activity, and as part of that mental activity, there was perception. That is to say, there was minimal consciousness and perceptual consciousness. If there is an inclination to doubt this, then consider the extraordinary sophistication of the activities successfully undertaken during the period of "unconsciousness".<sup>29</sup>

One way of making sense of this example, is to hold that there are two levels of consciousness; a lower one which is capable of discriminating between the road and the verge, the different colours of traffic lights, and is capable of making informed decisions based upon those discriminations - such as when to decelerate for an oncoming bend, and which road to choose to avoid low bridges. There is also a higher level of consciousness, that is aware of the lower level mental states and processes taking place. The means by which these two levels are related, may either be through the utilisation of concepts - referred to by Dretske as 'higher-order thought' (or HOT) theories - or via some quasi-perceptual scanner - referred to as 'higher-order experience' (or HOE) theories.

To make this clearer, consider the case of a person looking at a suitably illuminated cucumber. Common to both theories, is the acceptance of there being a low level representation of a green object; that is to say, the person will experience the greenness of the cucumber. However, whether or not the person is aware of the cucumber will depend upon either of one of the following. Firstly, some higher-level mechanism takes low-level data as input, processes it utilising the store of concepts available, and outputs a form of

---

<sup>27</sup> Dretske. 1993.

<sup>28</sup> Dretske. 1995.

<sup>29</sup> Armstrong. 1980. P 59.



awareness of that cucumber, related to how the input matches up with those concepts. Thus the level of awareness could range from being as specific as 'a *Cucumis sativus*', through 'a cucumber', 'a green vegetable', to a vague 'a thing just there'.

Alternatively, the higher-level mechanism involves an internal scanner, which is capable of being directed at the low-level states. Once scanned, a low-level state is brought into a state of awareness. The scanner is like a searchlight, that makes objects visible where otherwise they would be left in darkness. The crucial difference between these two mechanisms is that, *only* in the case of higher-order thought theories, are thoughts or judgements required to make a mental state conscious. This distinction becomes significant in the attempt to find a criterion for whether or not a creature has the faculty of consciousness. According to HOT theories, the creature must be able to use concepts; only HOE theories allow consciousness to exist where concepts are absent.

Dretske describes these as 'vertical' theories of consciousness. As will become clear, Tye and Rosenthal provide an account of consciousness that uses concepts to mediate between high and low levels, whereas Lycan employs the idea of an internal monitor. The following is a brief summary of these theories.

### Subsection 2 : Rosenthal's HOT theory of consciousness.

In "Thinking That One Thinks"<sup>30</sup>, our attention is drawn to two interesting points about the mind; that the term 'consciousness' is ambiguous, and that there is an assumption that all mental states are conscious states. The nature of the ambiguity emerges as follows. 'Consciousness' may be used either to describe a faculty of creatures, such that those in possession of this faculty are capable of instantiating conscious states - referred to as 'creature consciousness'; or it may be used to distinguish mental states of which we are aware, from those of which we are ignorant. It is the latter use of 'consciousness' - referred to as 'state consciousness' - in which Rosenthal is most interested.

The second point is that there is a danger of assuming that 'mental' is synonymous with 'conscious'. Admittedly, there are many states that are self-intimating, such as sensations of acute pain. However, it should be clear from Armstrong's example, as well as from reflection upon our states of mind, that we are not aware of every instance of thinking, believing or sensing. Indeed, the requirement that there *should* be a prior awareness of each mental state, would lead to an infinite regress<sup>31</sup>. This may be demonstrated with the following example. If I am in mental state 'A', then I am aware that I am in state 'A'. Being 'aware' is itself a mental state, so 'A' will be preceded by a

---

<sup>30</sup> Rosenthal. 1993.

<sup>31</sup> As pointed out by Ryle 1949.

state, 'B'. Since 'B' is a mental state, it too will be preceded by a further state of awareness 'C' - and so on ad infinitum. On this analysis, it would be impossible to reach the point where we could ever be conscious of anything. Instead we should admit that there will be mental states of which we are not conscious. This admission suggests that the real enigma is the mechanism by which we become conscious of mental states.

Rosenthal suggests that we become conscious of any mental state 'M' as the result of a higher-order thought, 'HT<sub>1</sub>', about 'M'. This thought need not be conscious; indeed, will only be conscious if we were to entertain a further higher-order thought about 'HT<sub>1</sub>'. This would occur when we introspect our thoughts and feelings, but otherwise, the feeling of immediacy of our consciousness is because we are not aware of 'HT<sub>1</sub>' but only 'M', the mental state it is about.

This contention is supported by a detailed consideration of the relationship between the 'expression' of a thought and the 'reporting' of a thought. Suppose I think that to-day is Tuesday. This thought may be expressed as follows: "It is Tuesday". It may also be reported thus: "I think it is Tuesday". Although these activities are distinct, it is deceptively easy to treat them as if they were equivalent; doubtless because any time I am in a position to assert 'that p', it would be appropriate to report that 'I am thinking that p'. Conversely, whenever I am in a position to make that report, then it is appropriate to make that assertion.

Yet this is to overlook two dissimilarities. First, whereas the expression 'that p', and the mental state 'p' have the same content, the contents of a report and the mental state reported are different; a report not only conveys the content of the mental state, but also the fact that the reporter is currently in that state. Second, the truth conditions affecting the expression of mental states are different from those of reports. This distinction can be demonstrated as follows. If a report were always true just when an expression were true, then the following would be contradictory: "It is Tuesday, but I do not think that it is Tuesday". Though the affirmation of an expression that 'p', and the denial of a report 'that p' is absurd, it is not a contradiction. This is taken to show that the 'expressing' and 'reporting' of thoughts are distinct activities. (This distinction holds true of other forms of expression also, such as 'gratitude' for instance).

Rosenthal points out that there is a close relation between saying something and thinking it; sincere speech is the expression of a thought. However, one may non-verbally express an unconscious thought, through one's body language or through Freudian slips for instance. Therefore, one need not be conscious of thinking in order to express a thought. This consequence is in contrast to the reporting of a thought. This activity conveys the agent's awareness of their thinking, as well as the content of the thought. In order to report a thought, one has to be conscious of that thought; this report is an expression of a

higher-order thought.

The argument for associating higher-order thoughts with consciousness of mental states proceeds as follows. The ability to report that one is in mental state 'M', is the same as the ability to express the thought 'T' - 'that one is in 'M''. Thus, 'T' is a higher-level thought. So 'M's being a conscious mental state coincides with having the ability to express the higher-level thought that one is in 'M'. If 'M' is not conscious, then we cannot report it; therefore we cannot express a higher-level thought about it. So when states are conscious, there will be a higher-level thought we can express. When states are unconscious, then there is no higher-level thought to be expressed. Therefore:

...we have every reason to conclude that a mental state's being conscious consists simply in its being accompanied by such a higher-order thought.<sup>32</sup>

Is the nature of consciousness so simple? The argument does not appear conclusive. The following is a brief rehearsal of the main points. We accept that it is possible to act, without being conscious of the thoughts that make those acts intentional rather than accidental. Therefore there is a mechanism whereby beliefs are linked with desires to produce behaviour, without our necessarily being aware of its operation. In this respect, such a mechanism is analogous to other physiological mechanisms that regulate other aspects of our behaviour; for instance, the processes by which we maintain body heat, blood pressure or insulin levels. In short, the mechanism by which thoughts are linked, is in some sense independent of our being aware of those thoughts. Without the property of consciousness to distract attention, it should be possible to formulate hypotheses on how these thought-processes operate. It is assumed that the higher-order thoughts involved in reporting, fall into the category of unconscious thoughts.

Consciousness arrives in the following manner. Awareness of the thought T, depends upon our having a higher-order thought T<sub>1</sub> concerning T: the difference between awareness and non-awareness, depends upon the presence or absence of a higher-order thought. This suggests that, when accounting for consciousness, all we need to explain is the presence of a higher-order thought. These types of thought, it has been suggested above, are in principle matters of knowable, physiological fact. Therefore, the nature of our awareness, reduces to the nature of relations between higher-order thoughts - and these are not very problematic at all.

The above may show that conscious states are accompanied by unconscious higher-order thoughts; but this falls a long way short of proving that we are aware of such states, in virtue of their being thus accompanied. The argument overlooks that fact that it begs the question to refer to components of these unconscious processes as

---

<sup>32</sup> Rosenthal. 1993. P 205.

'thoughts', 'beliefs' or 'purposes'. There is no guarantee that an 'unconscious thought' sufficiently resembles a 'conscious' one, in order to justify the assumption that it is only a matter of a state's relation to other mental states that makes them conscious or otherwise. Even if it were appropriate to refer to these unconscious processes in such terms, the following fact remains unexplained. Why is it that, not only thoughts, but all sensations, have the particular character that they possess? Simply adverting to a relation between mental states does not account for how incredibly dissimilar a conscious state is from an unconscious one: how could the conjunction of two benighted processes result in the refulgence of a conscious state? No analysis of the relation between 'expressing' and 'reporting' could dispel the queerness surrounding the idea. The mechanism by which we are conscious, remains as problematic as ever.

### Subsection 3 : Tye's HOT theory of consciousness.

In "Ten Problems of Consciousness", Michael Tye presents a version of Representationalism which may be regarded as another illustration of a HOT theory, although not included by Dretske as an instance of such. Such an omission is due partly to the fact that Tye's book had yet to be published, and partly because in Tye's 1992 Mind article "Naturalism and the Mental", he appears to accept the commonly held belief that some mental states - such as itches and pains: "...are not obviously representational at all."<sup>33</sup> However, between the time of the article and the book, Tye revised his position to include 'sensations' along with the more obviously representational states of 'perception' and 'thought'. The justification for holding that his position is HOT, lies in his contention that: "Phenomenal states lie at the interface of the non-conceptual and conceptual domains."<sup>34</sup> The following is a brief discussion on how he reaches this position.

Tye is interested primarily with what he calls lower-order states of 'Phenomenal consciousness'. These are the mental states that are characterised as having some sort of qualitative feel: they include, perceptual experiences, bodily sensations, emotions and moods. Separate from these, are higher-order states of consciousness, which may be involved in cognitive processes. These states only have a 'feel' property insofar as they are associated with one of the lower-order types of phenomenal consciousness. For example, the thought of my last birthday may be accompanied with a certain feeling. The feeling is there, not because of any intrinsic property of the thought, but because it is associated with a coincidental emotion of regret, for instance.

There are two reasons for assuming that lower-order states may be present even though the higher-ones be absent. Firstly because phenomenal consciousness is assumed to be present in animals

---

<sup>33</sup> Tye. 1992. P 431.

<sup>34</sup> Tye. 1996. P 144.

incapable of entertaining thoughts; and secondly, because of reasons alluded to, in explaining the different levels of awareness revealed by the truck driver example. A reason for accepting that the two levels are independent, emerges when one considers cases such as the Muller-Lyer illusion. Here, it is possible for the viewer to concede that one line *looks* longer than the other, even though they firmly *believe* the lines to be equal.

States of phenomenal consciousness are caused by the stimulation of our senses by objects both inside and outside the body. As such, the character of these states carries information about the associated stimulus. In an analogous way, because a growing tree renews its epithelial cells annually, so you can extract information from the rings of the tree concerning its age, and the growing conditions for any particular season. The idea is that the rings represent this information because their presence causally co-varies with the natural growth cycle. Tye asserts that the same is true of the senses. There is a relationship between properties of objects and the way they affect our senses. The character of our sensory experiences is caused by, and co-varies with, these properties. Therefore the character of these sensations carries information about those objects. So for instance, consciousness of orange represents the presence of an object within my field of view, that has a certain surface reflectance, and is illuminated in a certain way. The information contained by the character constitutes its intentional content, for it is linked to the objects of the world in virtue of having been caused by them.

Moods, emotions, and bodily sensations...are importantly like maps of our own internal physical workings, guides to our inner body states, graphic representations of what is going on inside (and to) our skins. Perceptual experiences are representations of the same sort, but their focus is on the outside world, the external terrain.<sup>35</sup>

According to Tye, phenomenal character thus described, has the following properties that enable it to interact with higher-order cognitive processes; these properties form the acronym 'PANIC'. The character is '**P**oised', inert as output of the sensory systems, in a state that is suitable for treatment as input to those processes that produce beliefs and desires. In order to be used as an input, the cognitive system has to be suitably attentive to the output processing and must possess the appropriate concepts. Any representation is '**A**bstract', insofar as it represents properties of objects in general, rather than of any particular object. They are '**N**on-conceptual' features that need not be covered by any concept that we could employ; that is to say, there will be an ability to discriminate amongst a number of different tasting substances beyond our conceptual labels for these tastes. It is '**I**ntentional' in the way already alluded to, and finally, it constitutes

---

<sup>35</sup> Tye. 1996. P 94.

the 'Content' of phenomenal character. It is a consequence of phenomenal character being a product of those properties to which we are sensitive, that we can only answer "What is it like?" questions that fall within the ambit of our sensory capabilities.

Tye subscribes to a 'Language of thought' type theory<sup>36</sup> concerning cognitive processes. However, one disadvantage with these theories, is that they are compatible with systems where conscious states are absent. It would appear therefore, that PANIC only reaches no further than the extent of our current knowledge of neurophysiology. As with Rosenthal's explanation, we are left with an unsatisfactory gap in our understanding. Tye stops at the point where there is an output from the senses. However, we are not conscious of the nature of this output until our cognitive system 'pays attention' to it. Here is where the gap lies, and it is not bridged by appealing to a 'cognitive system' and its activities.

#### Subsection 4 : Lycan's HOE theory of consciousness.

Lycan espouses a higher-order experience view:

...consciousness is the functioning of internal *attention mechanisms* directed at lower-order psychological states and events...attention mechanisms are devices that have the *job* of relaying and/or co-ordinating information about ongoing psychological events and processes.<sup>37</sup>

There are distinct advantages to accepting this account. Any theory concerning the nature of psychological states, must be able to differentiate between those of which we are conscious, and those which operate outwith our awareness. The above theory accomplishes this by assuming that the lower-level states are those which operate in the background, whilst our consciousness of them arises in virtue of their being attended to by some mechanism. We would also expect other creatures to have some form of awareness, though not necessarily the same as ours. The nature of their awareness would depend upon the sophistication of the attention mechanisms, as instantiated in different species. The mediation between mental state and consciousness by an attention mechanism, explains how it is possible that we may be more or less conscious of some things, and even deliberately suppress awareness of other states. Further, it accounts for the paradox of 'unfelt pain'; for instance, when we have a headache, our awareness of the pain need not be constant throughout the duration of ailment. The curious ability for pain to come and go, is explained by this combination of low-level states and monitoring processes; we are only aware of the pain when these mechanisms are directed upon the pain

---

<sup>36</sup> See Fodor 1975.

<sup>37</sup> Lycan. 1996. P 14.

states. Lycan stresses that he is not suggesting that there is a single mechanism, like an internal theatre, for that would be far too close to a Cartesian position. Rather, the idea is that there are a series of such mechanisms, each devoted to some aspect of the creature's welfare, which collectively contribute to its proper functioning.

**Subsection 5 : General problems with vertical theories of consciousness.**

Dretske, argues that all 'vertical' theories are inadequate. The fundamental problem with such theories is that, whilst talk of 'thinking of' or 'scanning' the lower-level states sounds initially quite appealing, it is difficult to convert the metaphor into a theory. The question, 'how do the higher levels become aware of the lower levels?' appears as problematic as the question, 'how do we become aware of the world?'. Unfortunately, that was the question the vertical theories were devised to answer.

Higher-order experience theories suggest that, for instance, a visual awareness of our surroundings, is the result of the monitoring of physical states associated with perception. However, it appears unlikely that the appearance of these physical states is going to be of any use, unless the information they contain is transformed into some other form of data. This qualification implies that there is something more than just 'scanning' involved. HOE theories are guilty of conflating 'vehicle' with 'content'. Where information is carried by a medium, it is essential to separate facts about the medium from facts about the information conveyed by that medium. For instance, the emotion of 'Despair' can be portrayed in the medium of stone; but one would not expect to understand the subject of a sculpture, by analysing the material from which it was carved. A scanning of the physical state would reveal its material properties, but one cannot guarantee that it would thereby reveal its psychological properties.

Even if it were possible to have a vehicle which was identical with its content, there remains the problem of explaining why scanning brain states should give rise to the sensuous variety of consciousness, and how this variety relates to the objects we believe we are sensing in the world. That is, it remains problematic how the world relates to the low-level experiences such that scanning these from a higher level should give rise to the conscious states of the senses of which we are finally aware. There is also a problem for being conscious of those mental states, such as thoughts, beliefs, which have no obvious felt quality. How is the scanner going to pick these up? That is, how may one become aware of beliefs, in a manner that is similar to the becoming aware of sensations?

Higher-order thought theories avoid the above problem by suggesting that we apply concepts to the lower-order experiences in order to bring them into consciousness. There are two objections that

Dretske makes against this theory though. Consciousness of any state requires that the creature must be able to manipulate concepts. There is evidence to support the notion that children are incapable of doing this for at least the first three years of their lives; animals also appear to have no facility with concepts. Therefore, if we adopt the higher order thought theory, we are committed to the idea that children beneath a certain age, and animals, are unconscious of their experiences. Even if we are prepared to accept this outrageous suggestion, there remains the problem of explaining how children acquire concepts from a position of unconsciousness. That is, how learning to manipulate concepts can be accomplished under conditions where consciousness is lacking.

The second objection adverts to the fact that our ability to differentiate between aspects of experience outruns the concepts we have for those experiences; the clearest example of this, is the number of shades of colour we can discriminate compared with the number of colour concepts that we hold. If a concept is required before we can be aware of a mental state, then we should be colour-blind to all those hues for which we lack concepts. The fact that we are aware of these hues, suggests that the link between concept and awareness is more complicated than higher-order thought theorists have realised.

#### Subsection 6 : Conclusion.

If the above criticisms are warranted, then Representationalism will have to look for an alternative interpretation of consciousness. The next chapter surveys the proposal that mental states are states in virtue of which we are consciousness, rather than states of which we are conscious - the view of the Horizontal theorist.



## Chapter Four

This chapter is devoted to a particular variation of Representationalism; the 'Horizontal' theory. This theory is the culmination of continuous refinement of a Materialist position, as outlined in the previous chapter. Certainly, there is much to say in favour of this view, but it is far from clear how much of it withstands scrutiny. The following four Sections, deal with the above theory in finer detail.

The first Section deals with the 'Horizontal' theory, as described by Dretske. This Section not only discusses problems associated with this position, but also suggests that it has a disturbing similarity to the 'Higher-order Thought' theories, that were dismissed in the previous chapter. In both cases, there is a strong conceptual element in the analysis of consciousness. The implications of this resemblance are not favourable to the success of a 'Horizontal' theory.

It is clear from Section One that Representationalism depends upon the validity of Externalism; yet there are reasons to doubt that Externalism may be so blindly trusted. The second Section will examine a range of arguments that demonstrate how little consensus there is over the Externalism-Internalism debate. Yet, whatever the outcome, there are reasons to suppose that the problem of inverted qualia has not gone away. In Section Three, these reasons are developed to show that the problem of qualia has yet to be addressed satisfactorily. Finally, in Section Four, a few concluding remarks will be addressed upon just how much further Representationalism has brought us, toward understanding the relationship between the body and the mind.

### Section One

Granted the difficulties with vertical theories, Dretske proposes what he calls a 'horizontal' theory, which eliminates all talk about levels of awareness. This Section will briefly outline his proposal, before assessing Dretske's particular version of Representationalism, as described in "Naturalizing the Mind". It will be suggested that this version fails to provide a coherent account of consciousness, and is possibly nearer a vertical theory than was intended. As such, it is vulnerable to all the objections levelled at the higher-order thought theories.

#### Subsection 1 : The Horizontal theory of consciousness.

The idea of levels of consciousness comes about through a failure to distinguish between states where 'we are aware of objects', and states that are 'our awareness of those objects'. Armstrong's example of the lorry driver has been taken to suggest that there is some property of the experience of seeing the traffic light for instance, of which we are may or may not be conscious.

Instead, Dretske suggests that it is a property of the experiencer that determines whether the experience is conscious or otherwise.

This contention is supported by the distinction between 'creature consciousness' and 'state consciousness' made by Rosenthal above<sup>1</sup>. The initial plausibility of the vertical theories comes from the following faulty reasoning: if one is conscious of hearing a ringing sound, then there is an experience of a ringing sound of which one is conscious. Instead of thinking that it is the individual by some mechanism that makes these states conscious, horizontal theories suggest that it is the nature of the states themselves that make the creature conscious:

Conscious mental states...are states we are conscious *with*, not states we are conscious *of*. They are states that make us conscious, not states that we make conscious by being conscious of them. They are states that enable us to see, hear, and feel, not states that we see, hear, or feel.<sup>2</sup>

Dretske argues that it is important to treat perceptual states as being determined by the properties of states of affairs, rather than of objects. Failure to do this, makes it difficult to appreciate that such things as dreams and hallucinations are veridical from the point of view of the experiencer. Consider the case of two individuals, both of whom report that they are looking at a woman's face. For whatever reasons, one is conscious of her features even though there is no object of perception; instead, there might be some complicated story to be told about memory and emotional triggering of the representation. The other is aware because they have normal eyesight, are facing her in suitable illumination, and have their eyes focused upon her. Both states are conscious, not because the individual is conscious of those states, but because there is a state that makes those individuals conscious of something. Our experiences therefore, are determined by the properties that experience represents something as having, rather than being determined simply by an experienced object.

The distinction between vertical and horizontal theories may be summed up as follows:

An experience of *x* is conscious, not because one is aware of the experience, or aware that one is having it, but because, being a certain sort of representation, it makes one aware of the properties (of *x*) and objects (*x* itself) of which it is a (sensory) representation.<sup>3</sup>

### Subsection 2 : Problems with the 'Internalist Intuition'.

Dretske is aware that his conclusions may strike some as an awkward fit upon the body of beliefs commonly associated with the mind. Accordingly, he provides an account of why this is so,

---

<sup>1</sup> Chapter Three. P 68.

<sup>2</sup> Dretske. 1995. P 100.

<sup>3</sup> Dretske. 1993. P 280.

accompanying his positive views. Many people are swayed by what Dretske calls 'The Internalist Intuition'. This 'intuition' holds that:

...experience (i.e., the quality of experience, what it is like to have the experience) supervenes on the constitution - and for materialists this can only mean physical constitution - of the experiencer.<sup>4</sup>

Hence, when two individuals are in identical internal physical states, then they are in identical mental states; and it is in virtue of their sharing those physical features that they may also be assumed to share the mental ones.

The Internalist intuition is taken to reflect a fact of life, rather than present a cogent argument. There are two factors which make this intuition less attractive. Firstly, it is an inadequate reflection of our mental concepts. Secondly, there is an alternative position - Externalism - that is compatible with much of what we consider to be essential to our notion of consciousness. For instance, we need not forego the belief that we enjoy some kind of special relationship with our mental states - a relationship that is not available to anyone but myself. (Burge argues for a notion of privilege that is none the less Externalist, in "Our Entitlement to Self-Knowledge", for instance.<sup>5</sup>).

Dretske opens his attack upon the Internalist intuition, by pointing out that it embodies an 'Act-object' view of experience. According to this view, visual experiences, such as seeing a yellow diamond for example, are the result of there being a yellow diamond in my field of vision, that causes a sense-datum in my mind. It is this mental image that I sense. This theory of perception clearly shows its Internalist sympathies: given that the sense-data are located in the mind (whether this is construed as a wholly separate substance from the physical, or whether it is identical in some sense with the physical), it would be impossible for two people to be the same internally and yet differ in what they sense. The most common objection to these views, is that they explain the process of perception at the expense of leaving the process of introspection quite mysterious; that is, the mechanism for sensing the mental image has yet to be explained. Whether or not Dretske gives a fair hearing to the 'Act-object' account, is arguable, yet so far as he is concerned, it is utterly untenable. Hence we should abandon forthwith any sympathy we may have for Internalism, that is prompted by its ability to sustain the above account.

Another reason for being more circumspect in our readiness to assume that identical internal states implies identical mental states, is demonstrated through what have become known as 'Replacement arguments'. These are thought experiments that were initially devised by Stich<sup>6</sup>, and were elaborated by Davidson<sup>7</sup>, before

---

<sup>4</sup> Dretske. 1995. P 145.

<sup>5</sup> Burge. 1995.

<sup>6</sup> Stich. 1983. Pp 165-6.

<sup>7</sup> Davidson. 1987. Pp 46-8. This thought experiment was introduced in Chapter Two above, Pp 32-33.

becoming the common property of Philosophers, to illustrate putative defects in Internalist theories. These arguments concern the ascription of consciousness to entities that have identical physiological states to ours, but who lack our evolutionary history. In this sense, they appeal to an Externalism about the processes that occurred to bring us to our present stage of development.

We are asked to imagine the spontaneous generation from particles in a swamp, of an entity which bears a physical resemblance to a human. Not only that, but it is also able to behave like an adult would. Granted a commitment to consciousness being created by a process of Natural Selection, we are not entitled to consider such an entity's movements as being "behaviour" in any human sense of the term, because the Swampman - hereafter S - is not the product of the same evolutionary history as humans. Noises are emitted from S, but not speech; it responds in the same publicly respectable way as any human, but its movements are entirely unmotivated: in effect, S is a Zombie.

The reason why we should refrain from ascribing mental states to S, is because of the Functionalist account of mental states that Dretske and others are observing. For a system to have a function, then it must have been designed to have that function. In the case of artefacts, it is the inventor who assigns the function to the system through the design. In the case of living things, it is the evolutionary history that - through natural selection - has imposed a de facto design upon systems within the organism, that qualifies those systems as having a function. In the case of a spontaneous creation of an entity resembling a human, there is no element of design. So by this account, there is no warrant for the ascription of systems, and hence no reason to believe that there is any degree of consciousness in such an entity. If we accept this analysis of S, then we have one less reason for finding Internalism an appealing theory, since it provides an example of an entity physically indistinguishable from ourselves, but which lacks consciousness. Therefore, the individuation of conscious states must depend upon more than mere physical identity.

### Subsection 3 : A defence of the 'Internalist Intuition'.

There are at least three reasons for failing to find the above considerations convincing. Firstly, the analysis of S's mental capabilities rests upon the assumption that, because he was designed neither naturally nor artificially, then his parts have no function. This is to misconstrue the relation between 'design' and 'function'. From the fact that an object was not designed to perform a particular function 'F', it is false to conclude that the object cannot be used to 'F'. What 'design' and 'function' have in common, is the relation that 'purpose' and 'function' bear to each other; and 'purpose' naturally suggests design or intention of some sort. It should be clear that if something has the purpose for 'A'ing, then it will have the function of 'A'ing. However, the entailment does not proceed in the opposite direction: it need not be the case that

something that functions as an 'A' was purposefully thought of as being an 'A'. How often does one resort to making-do with objects that have no particular purpose, or which have an entirely different purpose from the one being employed. One has only to think of instances of bubble-gum stopping leaks on radiators, coat-hangers as asexuals or flints as fire lighters. It just seems arbitrary to deny that these objects do not perform a function in some particular system, simply because they were not designed for that particular purpose.

The second reason for doubt arises from the difficulty in accepting that S would be unconscious; how would it be possible to pick him out from the common run of humanity? In the original example, S was created as a being physically identical to a person P, who was destroyed by the lightning at the same instant. Suppose P was reminded of a similar storm in his youth at the moment of destruction, then the brain state of S would be that of P having that memory. However, it is not a memory for S, so what sort of mental occurrence is it? Dretske wants us to conclude that it is not a mental state at all. This strikes me as being a little harsh on S. Certainly, his constitution disqualifies him from having memories of a life prior to the moment of creation; but that is just to say that the thoughts he has would not constitute memories, rather than show that he had no thoughts at all. If it were the case that for the mental state to exist, it must be correctly individuated by the agent, then it would appear to follow that simply because I have delusions of grandeur, it must mean that I am a Zombie. To mischaracterize a mental state, is not to deny the existence of any mental state - indeed, mischaracterization presupposes that there be a state, that one has incorrectly identified.

What appears to be going wrong is the identification of a process - 'remembering' - with the content of that process - a 'memory'. Certainly, there is no problem with describing S as 'remembering', if that refers to a mental process whereby data is retrieved from some sort of storage. What marks out P from the S is the way that the data came to occupy the space that it does; in the former case, through some method of processing occurrent experiences, and in the latter, through some freakish concatenation of electricity and matter. Therefore, I see no objection to describing the S as 'remembering', even though what he is conscious of is not a 'memory'.

The third reason is that there are other theories of teleology - such as the 'Relational' theory described by Walsh adverted to in the last chapter - which are compatible with the possibility of S's having sensations. Dretske has given no independent reason for preferring his account over these others, so his conclusions are by no means an inevitable consequence of accepting theories of Natural Selection.

#### Subsection 4 : Consciousness as a product of Evolution.

Regardless of whether or not the above considerations justify our dismissal of the Internalist Intuition, if Dretske can show his

Externalist view to be superior, then our feelings upon the matter will be irrelevant. It is to this view that we now turn. Dretske's analysis of consciousness relies upon both the involvement of Natural Selection, and the use of concepts. As will become apparent, these two factors do not sit easily with one another to form a unified account. The assessment of his position will begin by examining reasons for holding consciousness to be a product of Natural Selection; then his arguments to establish that the results of Externalism about 'thought' hold equally well for an Externalism about 'experience' will be considered. The main motivation for this account, is to provide an explanation of the nature of consciousness that is not constrained by internal states of affairs. Dretske is anxious to avoid the 'theatre' fallacy of consciousness. Vertical theories ask how it is that I become conscious of experiences; this gives rise to the problem of what is the nature of this ever-conscious 'I' which is the subject of these experiences. Horizontal theories hope to side-step this difficulty by explicating the notion of consciousness as something other than a relationship between two entities.

Dretske offers an analysis of experience in the following terms:

A state is a conscious experience of F...if the state has the natural (systemic) function of providing information about the F-ness of objects standing in the appropriate contextual relation (C) to the system.<sup>8</sup>

He states that it follows that consciousness is a product of Natural selection:

Since states acquire their systemic functions through an evolutionary process - here assumed to be natural selection - natural selection is being identified as the source and creator of conscious experience.<sup>9</sup>

The suggestion that a process such as Natural selection is responsible for the existence of states of consciousness, Dretske admits, may appear a little baffling. We are used to the idea that evolution does not create anything, but merely relates an organism's make-up to the environment in which the organism flourishes. If a feature appears in that organism, it is through some sort of mutation rather than some magic of the evolutionary process. For conscious states to play a rôle in our existence, we might assume they were present in our ancestors, and became developed over time; but what we would not think is that *because* of evolution we have consciousness.

However, Dretske suggests that this view fails to appreciate the rôle that Natural selection plays in his theory. It is not that 'consciousness' is selected as a trait, but rather, a separate characteristic is selected, which becomes 'consciousness'. To make this clearer, he offers an analogy between the way information

---

<sup>8</sup> Dretske. 1995. P 162.

<sup>9</sup> Dretske. 1995. P 162.

processing systems become conscious, and the way variable resistors become volume controls. Suppose someone were intent upon building an amplifier. They require a means of controlling volume. As there is no such thing as a 'volume control' per se, they purchase a variable resistor, which can be utilised to control the volume. These devices are not 'Volume controls' except when used in a particular context. They do not possess some nascent volume control-ness in their nature which is allowed to burgeon in the environment of an amplifier, but rather, they become volume controls in virtue of their being chosen to play that rôle.

The analogy with consciousness is quite suggestive. Just as there is no such thing as a 'Volume control' outwith there being a rôle for it to play in the context of an amplifier, so too is there no such thing as 'consciousness' outwith there being a rôle in the context of an organism. What plays the part of the amplifier, are the systems which provide the organism with information concerning the external and internal states of affairs, that have a bearing upon the survival of that organism. To provide this information, God goes into the parts shop and picks up bits of eye and nostril, to connect to the rest of the Central nervous system; and behold, the sensory systems produce states which represent survival-relevant aspects of the world. It is these informational states that are states of consciousness.

#### Subsection 5 : Problems with this account.

Imagine a world in which the processes of Natural selection did not take place. That Natural selection is a necessary prerequisite for consciousness, is supported by the contention that in virtue of his genesis, Swampman would lack consciousness. So, all creatures in that world would lack consciousness according to the above analysis. Yet it is hard to imagine that such a world could not have my twin, complete with conscious experience, even though the species to which he belonged did not develop by the process of evolution. Further, it might seem odd that one can attach such a strong philosophical claim to a contingent process: however, that does seem to be a consequence of this position.

Nor does this identification explain why conscious states were chosen as the medium for presenting information by Natural selection. Much of the body's systems are controlled by feed-back loops, in which consciousness plays no part whatsoever. Unless there is a perspicuous reason for preferring information to appear in one form rather than another, consciousness remains as mysterious as it has always done. One cannot suggest that more important information is conveyed by the senses, since this is not merely unsubstantiated favouritism, but relegates some of the more important physical systems to a second place that they do not deserve. It is demonstrable that the loss of many of the senses is compatible with staying alive, whereas it only takes the insulin levels to alter slightly, and the result is death.

Further questions might arise as to the consequence of basing the development of consciousness as a result of a process. This would imply that any organisms that develop through Natural selection, should have conscious states of their environment. This is - perhaps - acceptable as it stands; but when it is linked with Dretske's Externalist position regarding the conceptual element of consciousness, then this implies that all organisms must be able to apply a range of concepts to their sensations. This consequence is a little hard to accept. If consciousness is to be the sole preserve of a range of creatures, then there needs to be an elucidation of which features of the process of evolution could make this so. Given that all life has much the same needs, there has to be some reason why humans, at least, are so distinctive in their perceptual abilities; and this reason cannot be furnished simply by adverting to a process which it is suggested that all living things undergo.

#### Subsection 6 : The Externalist element in Horizontal theories.

The arguments of Putnam and Burge appear most plausible when applied to concepts, and by extension, to the contents of propositional attitudes. However, there is a recognised difference between those mental states to do with cognition, and those to do with sensation. What is not immediately clear, is why we should accept an Externalist account of sensations; these seem to be more clearly dependent upon the nature of our internal constitution, rather than our environmental situation. Dretske makes our awareness of sensations dependent upon our having the appropriate concepts. For any sensible property F and any subject S, Dretske asserts that:

Without the concept of F, S is "blind" to the F-aspect of his phenomenal experience (though not, of course, to the external objects this is an experience of). He cannot be made conscious of it.<sup>10</sup>

Hence, if we cannot be said to individuate experiences without the aid of concepts, and concepts are determined only by external considerations, then our experiences are affected by external considerations.

Dretske offers a choice concerning the nature of phenomenal experiences. Either they may be identified with thought-like entities, such as potential beliefs, suppressed inclinations to believe or micro-judgements; or phenomenal experiences are distinct from thoughts. In the first case it follows trivially that, insofar as thoughts are externally individuated, then so too will the thought-like entities of experience. In the second case, if one is able to separate the experience that 'something appears in a certain way', from the belief that 'it appears that way', then there will be qualia of which one is completely unaware.

---

<sup>10</sup> Dretske. 1995. Pp 134-5.



Dretske concludes:

If one takes qualitative states as essentially knowable...then phenomenal experiences must be externally grounded if beliefs are. If, on the other hand, one is willing to tolerate unknowable qualia, what reason is there to insist that qualia must be *the same* in physically identical beings? <sup>11</sup>

Hence, one should regard phenomenal experiences as being identifiable with thought-like entities.

Dretske describes two cases which illustrate how experience is dependent upon concepts. The first case provides a negative proof that Internalist criteria do not correctly individuate experiences. The second case provides positive proof that without concepts we would lack awareness. In doing so he employs the distinction between 'phenomenal' and 'doxastic' types of experience (signified by a subscript 'p' or 'd'). The description, "the top line looks<sub>p</sub> longer than the bottom line" refers to my visual experience, even though the lines actually look<sub>d</sub> the same length. 'Doxastic' experiences constitute the norm against which one discriminates amongst veridical and non-veridical perceptions.

Schematically, the first case involves an object 'O' which looks<sub>p</sub> 'F' to an individual 'S'. S has the concept F, and is therefore able to believe that O looks<sub>p</sub> F. However, O looks<sub>d</sub> G and not F.

An examiner briefly holds up seven fingers. Once the fingers are out of the field of vision, the examiner asks the subject how many fingers were held up. Suppose the subject replies "Eight", because that is the number they thought they saw. Though what was present to their eyes were seven fingers, what they thought they saw presented were eight fingers. They would hold the same belief as one who thought they saw eight fingers when they were shown eight, even though they were in a different visual state from the person who saw eight fingers.

In order to make sense of their reports, we have to assume that each person saw eight fingers: that is, both experienced the sight of eight fingers being held up. So their experiences in that respect must be the same. However, the two cases are importantly different, since in the first case there were only seven fingers held up. The conclusion is that there are facts about our phenomenal experiences of which we are not always aware. Further, in order to determine whether the subject actually saw seven or eight fingers, there is no point referring to their beliefs, for these were the same - that they saw eight fingers - but rather to the outside world where the fingers were. What this is taken to demonstrate, is that we can only discriminate between the two subject's experiences by reference to the environment, and not the belief states of the subjects. So Externalism is preferable to Internalism.

In this example, there seems to be an equivocation between "what is seen" as characterised by some description, and "what is seen" as characterised as some process carried out by the visual system. It is true that we can be more or less scrupulous in our

---

<sup>11</sup> Dretske. 1995. P 141.

attending to what is available to us visually; we abruptly become aware of features in our field of vision, not because they have suddenly materialised, but because we had not bothered to look carefully enough. Absence of features of our visual field may be caused by a lack of attention, and these are quite different from the absences from our vision that are posed by lacunae in our vocabulary. In effect, the example above is not straight-forwardly about vision, as about memory - what we can remember about what we saw. It is not clear that this example demonstrates that external criteria determine the content of a propositional attitudes, rather than determine the character of experience.

Schematically, the second case involves an object *O* for which *S* lacks the relevant concept. Hence, *O* appears<sub>p</sub> *G* to *S* without it appearing<sub>a</sub> *G* to *S*.

Consider a subject *A*, listening to a piece of music. Grant that *A* has an extensive theoretical knowledge of music, and perfect hearing. Unless they know what a change of key sounds like, they can listen to the music without being able to detect the occasions when the key changes. Alternatively, a different individual *B*, hearing the same sounds, and thereby experiencing the same noises, may be able to pick out every change of key as they occur. *A* and *B* have not heard different things, and yet the two people are different with respect to the thoughts they have about what they hear. *A* fails to hear changes of key, whereas *B* is successful in picking them out. As an experience, the key change does not exist for *A*; not because of some physiological defect, but because of a conceptual one - *A* is 'deaf' to key changes. If concepts play this essential rôle in experience, and if concepts are externally individuated, then the external world will be essential in individuation of experiences.

#### Subsection 7 : Problems with this analysis of consciousness.

This again seems to be confusing the discriminatory abilities of language on the one hand, and our senses on the other; that is, a conflation of being aware *that* 'x', and being aware *of* 'x'. Though I may be deaf to key changes, it surely is not to be suggested that I lack any form of auditory experience whatsoever. Yet this is denied by Dretske:

I do not need the concept RED to see red. But I do need this concept to become aware of the *quale* red, to become aware that I am having an experience of this sort. This being so, qualia...remain "hidden", inaccessible, until one acquires the conceptual resources for becoming aware of them.<sup>12</sup>

However little I am able to make of the stimulation of my senses, by way of being able to classify them as being of one sort or another, it does not mean that I lack those experiences. That this should be clear, would result from a moment's reflection upon the consequence of assuming that our concepts are prior to our experiences. For this ranking makes it utterly mysterious how we

---

<sup>12</sup> Dretske. 1995. Pp 139-40.

acquire concepts before we have experiences. In order to learn a language, it must be communicated to me in some manner. It must therefore be possible for me to have some means of discerning this communication. How this is to be so, given my initial concept-free zombie-like state, is inexplicable, except were one to resort to something like a notion of innate ideas. But such a notion would be of little use to an Externalist, who wants to keep meanings outside the head.

There is a further difficulty. On this account, one can have experiences of red, even though one is unaware of them when one lacks the concept 'red'. But what is the nature of these pre-conceptual 'experiences'? Either our senses produce a qualitative output independently of our concept possession, or they do not. If they do, then regardless of our awareness, it still remains to be explained how this unconscious phenomenal character can be Naturalised. If they do not, then the mechanism by which concepts can produce phenomenal character out of non-phenomenal activity needs to be explained.

Admittedly, this is not a vertical theory, in the sense of suggesting a relation between an experiencer and experience, mediated either by concepts, or some form of scanner. However, it is clear that the theory makes the possession of concepts a prerequisite for awareness of our mental states. In this sense, this account shares a number of features with Rosenthal and Tye. At least one criticism of HOT theories is based, not upon anything distinctly vertical about concepts, but rather, that they entail that creatures lacking concepts must thereby lack experience<sup>13</sup>. This would have the absurd consequence of denying consciousness to infants and animals. It is unclear how the involvement of concepts in the Horizontal theory, is able to avoid the force of these criticisms.

As Stich noted above<sup>14</sup>, it is notoriously difficult to provide the necessary and sufficient conditions required to determine whether or not a particular object falls under a concept. If we accept this analysis of concepts, then it has the unfavourable consequence of making our states of awareness indeterminate, whenever such conditions cannot be established. If the notion of 'recognitional concepts' is used, then the distinction between HOT and HOE theories collapses, insofar as the 'scanner' could be interpreted as exercising recognitional abilities. This objection would have no force against theories like Lycan's, though that may be small comfort in the light of other criticisms to which it is vulnerable.

Finally, there is a general problem concerning the association between concepts and experiences. If experiences are dependent upon concepts, and if experience is a product of our evolutionary history, then there seems to be something of a competition between concepts and evolution as to who should have the palm for providing us with consciousness. On the one hand, the process of Natural selection has been acknowledged as the creator of consciousness,

---

<sup>13</sup> Dretske. 1995. Pp 110-112.

<sup>14</sup> Chapter Three, P 62.

but on the other hand, it is stipulated that qualia remain “hidden” and inaccessible until we have the conceptual resources to become aware of them. There is no doubt that Dretske requires this conceptual constraint if he is to get from being an Externalist with respect to thoughts, to being an Externalist with respect to experience; but this desire will remain unfulfilled whilst he holds that consciousness springs from Evolution; for between the contingency of Natural selection and the necessity of concepts, there seem to be two incompatible contenders required for the presence of consciousness. (It is worth noting that in “Is Experiencing Just Representing?”<sup>15</sup>, Block argues for just such a conflict. Briefly, he suggests that some interpretations of Natural Selection are at odds with our intuitions concerning Swampman, or they run into problems concerning the relation of phenomenal character and intentional content.)

## Section Two

Representationalism is regarded as being consonant with Externalism; indeed, Block presents an argument which suggests that a coupling of Representationalism and Individualism, is a marriage that would barely last long enough to allow the confetti to settle<sup>16</sup>. Whether or not Externalism is forced upon them, it is nonetheless clear that most accounts of Representationalism regard this association as reflecting favourably upon their thesis. The debate between Externalism and Individualism is too vast. The limited intention of this discussion, is to demonstrate that the validity of Externalism is far from obvious; thus, any sympathy we have for Representationalism must be independent from our assuming that Externalism is correct. In what follows, the terms of the debate are introduced; after which, a sample of three articles are presented, which dispute the meaningfulness of the distinction between the two sides. The principle of Realizability<sup>17</sup> is taken as holding that mental state types are instantiated by physical state types, be they construed as physical properties or functions. A mental state is said to be realized physically whenever such an instantiation occurs. The final subsection is presented to show that the principle of Realizability is undermined by an Externalist interpretation of mental states.

### Subsection 1 : The origin of the conflict between Externalism and Internalism.

The debate concerning Externalism and Individualism centres upon the relationship between thoughts and physical states, and how they can be specified in order to be utilised by psychological theories. If you are inclined towards being a Materialist of some

---

<sup>15</sup> Block. 1998.

<sup>16</sup> See Block. 1996.

<sup>17</sup> To avoid confusion, whenever this principle is being adverted to, ‘realize’ and its cognates will be spelt with a ‘z’.

sort, then there are at least two intuitions that spring to mind about this relationship. One is that such thoughts as we have, are caused by some sort of interaction between our senses and our environment. So, were I to direct my gaze toward the Scott Monument, and think about what is in my field of vision, the thoughts I have could be appropriately described as "Thoughts about the Scott Monument". Were it not for the action of that building upon various neurophysiological processes, then I would not be having that particular thought. If it were a different building, then I would have a different thought; so thoughts may reasonably be classified according to a specification of the environment.

The second intuition is that each time I think something, there is a process in the brain that instantiates that thought; indeed, the only means I have of thinking, is because my brain is capable of performing such tasks. Such is the link between the brain process and the resulting thought, that where there are two different processes, there will be two different thoughts; so thoughts may reasonably be classified according to a specification of the brain process.

What brings these two intuitions into conflict, is the possibility of two physically identical individuals occupying different environments. In such a situation, the above mentioned principles of classification are in competition; according to the first, such thoughts would count as being different, because the environments were different; according to the second, the thoughts would count as being the same, because the physical state of the individuals was the same. If Materialism is to be a viable option, then it has to resolve this clash of intuitions.

Attempts at resolving this problem have resulted in two identifiable positions. 'Externalism' is the thesis that classification of mental states should be made according to factors in an individual's environment, such that the content of an individual's thought reflects the specification of their environment. Alternatively, 'Individualism' is the thesis that classification of mental states should be made according to the specification of an individual's physical state. Due to a series of articles by Putnam and Burge, many accept the Externalist position as the correct way of classifying thoughts. It is not the purpose of this discussion to challenge Putnam's semantic thesis that "meanings ain't in the head": rather, it is to cast doubt on the assertion that in order to provide an explanatory taxonomy of thoughts, we *have* to be Externalists.

The following is a brief description of the influential "twin-earth" thought experiment. In common with a majority of descriptions of this argument, 'water' will be used (It ignores the fact that 'water' is an unhappy choice of substance to employ in this case, since it raises problems about the chemical identity of the twins: but somehow, alternatives lack the attractive patina that years of philosophic usage have given to 'water'.). Consider a normal English-speaking member of the human race - call him 'Ed'. Assume that although he does not know its chemical composition, he

is otherwise considered a competent user of the word 'water'. Now consider the possibility of a twin-earth, which is identical with earth, except that there is no  $H_2O$ . Instead, there is a substance with all the properties of  $H_2O$ , referred to as 'water', but with the chemical composition XYZ. On twin-earth, there is a physiological double of Ed, called Ted. Just like Ed, he is unaware of the chemical composition of the stuff he calls 'water', but he too is judged to be a competent user of the twin-English. Ed is lodging at a small bed and breakfast in Tobermory, and asserts ingratiatingly to his landlady that "The water on Mull produces a better Malt whisky than that on Speyside". Ted in the twin-earth equivalent to Tobermory, is making the same assertion to an equally unimpressed twin-landlady. However, whilst they are by hypothesis physiological doubles, they are never the less thinking about  $H_2O$  and XYZ respectively. That is to say, sameness of physiology does not imply sameness of thought.

The situation becomes unsettling when Ed is unknowingly transported to twin-earth. The  $H_2O$  thought he had on earth, is now a distinct XYZ thought on twin-earth, in spite of the fact that he is unaware of such a difference. This means that any attempt to explain Ed's behaviour on twin-earth, will have to attribute thoughts to him that he is incapable of having - that is, he has  $H_2O$  thoughts even though there is no  $H_2O$  to think about, or he is having XYZ thoughts even though he has absolutely no idea what XYZ is. The concern is that Externalism here undermines our explanatory abilities.

Subsection 2 : Egan - there is no useful distinction between these positions.

The first article I wish to consider, is "Must Psychology be Individualistic?" by Frances Egan. The purpose of her argument, is to show that psychological types as used in explanations of behaviour, may be either Individualistic or Externalistic according to the purposes of the theory within which the types appear. To do this, she examines the arguments of Fodor and Burge, and finds that neither philosopher is able to establish their conclusions. Either they assume, without argument, that certain supervenience relations hold; or it is a consequence of their positions that such disciplines as Geology, Biology and Linguistics for instance, could not be regarded as scientific.

Egan's comments upon Individualism are directed toward the view expressed in "Psychosemantics" by Jerry Fodor<sup>18</sup>. Here, Fodor suggests that science individuates types according to causal powers; that is to say, according to the intrinsic properties of an entity. This is meant to establish that no taxonomy could be based upon external or extrinsic properties, since these have no bearing upon an entity's causal relations. Furthermore, he claims that we would be unable to make any sense of mental causation unless we insist that mental states supervene locally upon brain states. Egan contends that neither of these claims establish Fodor's position.

---

<sup>18</sup> Fodor. 1987.

Against Fodor's initial claim, she provides an argument analogous to his, showing that there are some cases where sameness of physical constitution, and therefore of causal powers, would not lead us to conclude that there was a sameness of content. Her example involves the case of an answering machine, supplied with an identical message, which would serve two addresses. The message is ambiguous between "I love you", and "Isle of view". Though the messages are produced by the same type of mechanism, they will produce distinct responses according to their use by a Dating service or an hotel. In this case, we are happy to say that sameness of mechanism does not imply sameness of content. From this example, Egan derives the moral that too often we consider these thought experiments with a bias toward certain supervenience relations. As the twin-earth example shows, at first glance it is equally plausible that behaviour supervenes upon the intrinsic nature of brain states, as upon the extrinsic nature of mental states. All Fodor's argument shows is that it is plausible to individuate according to causal properties, not that one must.

Fodor's second claim, that only Individualism makes sense of the relation between the mind and the body, is also suspect. Egan observes that Fodor is interpreting the Externalists as saying that the environment has a causal influence upon the individuation of psychological types. Instead, their contention is that the environment is relevant to this process. Many sciences classify according to criteria that are not based upon the physical structure of the object; for example, Geology classifies according to the origins of the rock - be they igneous, sedimentary or metamorphic. Biology and Linguistics are further examples in this vein. There appears to be no justification for judging these forms of classification as unscientific, simply because they advert to extrinsic features of these objects rather than their micro-structure. In which case Fodor cannot simply say that, if psychology is to be a science, then it must classify its types according to their causal properties.

Against Externalism, Egan comments upon two claims made by Tyler Burge in "Individualism and Psychology"<sup>19</sup>, that purport to show that some psychological practises are non-individualistic. She denies that these arguments support Burge's conclusion. The first claim is developed via a thought experiment, similar in kind to Putnam's twin earth example. It concerns two individuals called Bert, physically identical, but inhabiting different linguistic communities. Bert<sub>1</sub> believes he has arthritis in his thigh, even though this is a malapropism - arthritis, as every philosopher of mind knows by now, being a disease of the joints. Similarly, Bert<sub>2</sub> believes he has arthritis in his thigh; however, Bert<sub>2</sub> belongs to a community where the term 'arthritis' covers diseases of the joints and upper leg. Therefore, the two Berts exemplify an instance where individuating psychological states according to physical types would be insensitive to an important difference between their beliefs - that

---

<sup>19</sup> Burge. 1986.

is, only Bert<sub>2</sub> believes correctly that he is suffering from the disease.

Psychological theories refer to an individual's mental states via oblique attitude ascription. The above thought experiment demonstrates that the content of these oblique attitude ascriptions does not supervene upon the physical state of the agent: even though their physical states were the same, the content of the Berts' thoughts was different. Therefore, Psychologists are Externalists when it comes to referring to the contents of propositional attitudes. However, Egan points out that Burge's argument works only if one accepts that psychological theories treat sameness of ascription as implying sameness of psychological content. She cites the work of Brian Loar<sup>20</sup> upon this assumption; he argues that sameness is neither necessary nor sufficient.

Loar gives two Burge-style instances to support his argument. The first shows that sameness of oblique attitude ascription is not necessary. Two people's beliefs are regarded as being the same by a psychological theory, even though their beliefs are distinguishable by oblique attitude ascription. He uses a twin-earth situation, where a psychologist is in possession of a patient's diary, but does not know to which earth it refers. An entry reports the diarist's decision not to go swimming because the "water" was too rough. The psychologist does not know whether water or twater is being referred to, so cannot ascribe to the patient the appropriate belief content. However, Loar points out that this would not prevent the psychologist from providing a satisfactory explanation for the diarist's behaviour; the diarist did not want to swim, because they believed the stormy water was dangerous. Whilst there is nothing to prevent separate theories being applied to each possible belief - be it twater or water that was the subject of the diarist's concern - no advantage would be gained by differentiating between these beliefs. After all, there is nothing significantly different between water and twin-water that would account for differences in behaviour (If you were afraid of water because you might drown, then you would have every reason to be afraid of twater). Rather, an important generalisation would be lost that associated one type of behaviour with several distinct substances.

The second example shows that sameness of oblique attitude ascription is not even sufficient; here, even though an individual's behaviour can be explained by the affect of two of his beliefs, the difference between these beliefs does not emerge through analysing them in terms of oblique attitude ascription. He adapts Burge's 'arthritis' example. A person, 'P', believes they are suffering from arthritis. They get into conversation with a Frenchman, who tells them about a condition called 'arthrite'. 'P' realises that he is suffering from 'arthrite', but does not realise that 'arthrite' is the French word for arthritis. Loar suggests that, from the psychological point of view, it is possible to distinguish his behaviour with respect to his arthritis and his arthrite, even though these beliefs would be the same as analysed as oblique attitude ascriptions.

---

<sup>20</sup> Loar. 1988.



Generally, it is vital to note that psychological theories individuate beliefs according to the rôle they play in conjunction with other beliefs, stimuli and behavioural dispositions – what Loar refers to as their ‘conceptual, or functional, rôle’. What is clear from the above examples, is that the individuation of a set of beliefs according to oblique attitude ascription, will not always coincide with that according to their conceptual rôle. Egan comments thus on the relevance of Loar’s work in clarifying the issues in this debate:

Loar’s examples suggest the possibility that psychological theory may individuate mental states individualistically even though the language it uses to refer to these states is not itself individualistic.<sup>21</sup>

The second claim Burge makes in support of Externalism, has to do with perceptual states. Therefore, it is of especial significance to the cogency of Representationalism, given that perception is regarded as a way of dealing with the problem of intentionality. Burge employs David Marr’s<sup>22</sup> theory of vision to establish his point that perceptual states may be individuated according to features of the environment. Although there is a correspondence between Egan and Shapiro<sup>23</sup> upon the extent to which Marr’s theory is individualistic, which is far from conclusive as yet, the fact of their dispute underlines the contention of this section; the issue between Individualists and Externalists is far from settled.

### Subsection 3 : Walsh - resolving the conflict by appeal to context.

The second article that undermines the resilience of the Externalist position, may be found in an article by Denis Walsh. In “Wide Content Individualism”<sup>24</sup>, he argues for a way of reconciling our intuitions upon the question of how to individuate psychological kinds. Walsh notes that there has been a fundamental misunderstanding between the rivals, concerning what each camp claims. He cites Burge as providing an example of how the Externalists see the Individualist view<sup>25</sup>, and notes that Burge assumes that Individualists deny the environment any rôle in the individuation of psychological states. Yet the Individualists are equally at fault for assuming that when a physical state realizes a particular psychological type in any context, then it realizes that type in all contexts. It is Walsh’s view that the individuation of psychological kinds depends upon their teleological function. Further, it makes no sense to identify the function of a trait, without placing that identification within a context. That is to say, individuation of psychological types is context sensitive; so where context is shared, sameness of physiological state implies sameness

---

<sup>21</sup> Egan. 1991. P 195.

<sup>22</sup> Marr. 1982.

<sup>23</sup> See Egan 1992 and Shapiro 1993.

<sup>24</sup> Walsh. 1998.

<sup>25</sup> See Burge’s definition of Individualism in Burge 1986, Pp 3-4.

of psychological state; where context is different, no such implication exists.

He provides a definition of context thus:

...an individual is *in* a context,  $\gamma$ , only if she is causally connected in the appropriate way to an environment characterised by  $\gamma$ .<sup>26</sup>

Here, context  $\gamma$  is determined either by its historical relation with the individual, as may be the case in a stable evolutionary environment; or by its occurrent situation, such as when an abrupt change has brought about a new trait; or by the linguistic community in which the individuals find themselves. What lies behind the debate, is a failure to account for the relevance context has in the practice of psychological explanation. The above definition of context enshrines the Externalist view that the environment is relevant to how we mark out psychological kinds, but it retains the intuition that causal connections are also important in such activities. In those cases where our intuitions find both theories unsatisfactory, such as the Twin world instance outlined above, it is because the description of the contexts has been omitted from the example. Hence, there is no requirement to adhere to one theory at the expense of the other; the appeal of both can be retained by supplementation with this notion of 'context': hence a philosophical position cannot be dismissed or accepted simply because it is Externalist or Individualist.

It is interesting to note that Representationalists such as Dretske and Tye, make use of 'context' in their theories. In their case, it is used to distinguish between two phenomenally similar experiences, such as when I see a book, and when I hallucinate a book. There is nothing in the representation itself which will differentiate between these two types of experience. Rather, it is the context of the experiences which is where the difference lies, and in virtue of which one is able to speak of misrepresentation.

#### Subsection 4 : Kirk - problems with Ultra-Externalism.

The final article to be considered, written by Robert Kirk, concerns the limit to which Externalism makes sense of sensations. A lot of the work done in this debate concerns the propositional attitudes, especially beliefs, at the expense of sensory experience. In "The Trouble With Ultra-Externalism"<sup>27</sup>, Kirk redresses this balance, by concluding that Externalism goes too far in its individuation of perceptual states. He provides a thought experiment to support this claim. An individual who has been blind from birth, is provided with an apparatus for discriminating between two colours of light. This device consists of a light sensor with two outputs; one which produces a signal type 'G' whenever it is in the presence of green light; and the other, a signal type 'R', whenever it

---

<sup>26</sup> Walsh. 1998. P 642.

<sup>27</sup> Kirk. 1994b.

is in the presence of red light. Each output is connected to a separate part of the brain; 'G' signals to region 'A', and 'R' signals to region 'B'. In time, the patient learns to associate one particular experience with 'G' signals, and a different experience with 'R' signals, and thereby is able to tell when there are green or red lights within the range of his sensor.

Kirk argues that the quality of the experience is not determined by the environment. Although different parts of the brain produce different experiences when stimulated, this has nothing to do with the environmental cause of the stimulation: for as he points out, the 'G' type signal could have been attached to region 'B' and thereby set up an association with a different experience. It seems instead that the experiences are best typed according to internal states rather than according to what is happening in the patient's environment. That is to say, it appears from this thought experiment that perceptual states might well be more reasonably associated with an Individualistic rather than Externalistic account of individuation.

Again, there is much more that should be said on this matter, and this argument will reappear in Section Three, when the question of spectrum inversion is addressed. However, it serves now to underline the wariness that should be employed before wholeheartedly embracing Externalism.

#### Subsection 5 : Is Externalism consistent with Representationalism?

Lastly, the following is intended as an argument to show that the Externalism in Representationalism, is at odds with its commitment to causal co-variation. Whatever the ultimate nature of matter is discovered to be, this argument assumes that matter is consistently law-like in its behaviour, and that everything is composed of it. The debate between 'Individualism' and 'Externalism' recognises a boundary between the physical stuff of the individual and the physical stuff of the rest of the world. For convenience, I shall refer to these spaces as 'the brain' and 'the environment' respectively.

Twin-earth situations exploit this separation, by assuming that not all properties of a specified environment will causally interact with the brain, so that there can be changes in the environment which will not cause any changes in the brain. This assumption is wholeheartedly accepted. However, such experiments go on to assume that it is possible to have two different mental states without there being a change in brain state. It is this assumption that is being challenged. The argument is similar to that put forward by Fodor in "A Modal Argument for Narrow content."<sup>28</sup>

The twin-earth thought experiments suggest the following. Firstly, that distinct thoughts can be attributed to individuals with the same brain state - therefore there is no definitive classifying relation between the brain state and the mental state. Secondly, the same brain state can be attributed to an individual, even though they

---

<sup>28</sup> Fodor. 1987.

are transported to a different environment - therefore brain states are unaffected by at least some changes in the environment. Thirdly, mental states are affected by the environment, such that the mental state of an individual in one environment, need not be the same as when that individual is transported to a different environment - therefore, there is a definitive classifying relation between the environment and the mental state. Finally, there will be some physical changes in the environment which will have no effect upon the brain state, but which are nevertheless reflected by a change in the mental state.

The problem is, how does a Representationalist explain a mental change which has no corresponding physical change, if they are committed to the idea that mental states are realized by brain states that causally co-vary with the environment? Fodor's conclusion is that you cannot provide such an explanation; that any individuation of mental states that does not classify contents of thoughts in terms of the brain state, is required to employ some rather fancy causal story - such as, the environment causes thoughts to change whilst by-passing the brain.

One response to this line of argument has been raised by Frances Egan, in "Must Psychology be Individualistic?"<sup>29</sup>. Egan points out that the above considerations have misconstrued the Externalist position. Rather than saying that the environment *causes* the difference between the earthian and twin-earthian thoughts, the point is that the environment is the basis for picking out one thought from another. So the debate is at cross-purposes, with Fodor talking about the mechanism whereby one comes to have the thoughts that one does, and the Externalists talking about how to classify those thoughts, regardless of how they are realized. All that a Representationalist is committed to, is the belief that mental states are realized as brain states, not that the same brain state can realize one and one only mental state. Though mental states are realized physically, it need not be the case that the same physical state type  $P_1$  can only realize a single mental state type  $M_1$ ; in the twin-earth example,  $P_1$  is both a realization of  $H_2O$  water thoughts and of XYZ water thoughts.

The following is an argument that attempts to deny the cogency of Egan's response. The twin-earth example set up as follows. The physical differences of the environment have no distinctive causal properties with respect to the brains of the twins. The assertion that the individuals have different thoughts, is backed-up by two claims: that there are environmental differences between the two worlds; and that the twins' brain states are capable of realizing different thoughts, even though they are physically identical. It is this latter claim that requires some examination.

If there is to be an assimilation of the mental to the physical, then the notion of 'realization' of the mental by the physical should be understood in virtue of the properties of physical matter. Merely correlating mental and physical states would not be enough to

---

<sup>29</sup> Egan. 1991.

underpin a scientifically respectable classification. The reason for asserting that mental type  $M_1$  is realized by physical type  $P_1$ , is based upon an understanding of the properties of  $P_1$ , rather than the bare fact that the two events happened to coincide. Any serious attempt to identify mental states with physical states or processes, must observe the principle, that any discernible difference at the mental level should be matched by a discernible difference at the physical level.

The identification of those physical types that realize mental types, must assume that realization is consistent and principled. Otherwise, there would be no justification for saying of any passage of brain activity  $B$ , that it was the realization of a particular thought  $T$ , rather than a different thought  $T'$ . Most philosophers accept that a single mental type  $M_1$ , can be realized by distinct physical types  $P_1$  through  $P_n$ . It is no more surprising than accepting that distinct physical processes can all perform the same function; so for instance, there are several distinct ways of opening a can. However, there has to be some feature that  $P_1$  through  $P_n$  have in common, for justifying their being singled-out amongst other physical types, as the instantiation of  $M_1$  - this could include a functional similarity, in some abstract sense. Without this stipulation, there is a danger that the association of  $M_1$  with any of  $P_1$  through  $P_n$  becomes arbitrary.

Just as associating mental type  $M$  with brain type  $B$  only because the two happen to correlate, does not advance us in understanding how mental types are realized by brain types, similarly neither does the association of environmental type  $E$  with brain type  $B$  only because the two happen to correlate, provide any greater understanding of this relation.

To employ the vocabulary of supervenience, the realization principle regards the mind as supervening upon the subvenient properties of the body. Thus, any change in the mental is matched by a change in the body. Clearly, this type of supervenience is not being used in the twin-earth case, where there is a change in the mental that is not reflected in the body. Rather, the mind is regarded as supervening upon the state of the environment. That is, the environment becomes the subvenient base; any change in the mind is reflected by a change in the environment. This is acceptable to some<sup>30</sup>, but there is a hidden cost.

Representationalism assumes that the content of phenomenal character is a result of causal co-variation of events in the world and our phenomenal states. This is analogous to the relation between the environmental factors causing variation in growth rings of trees. If there is no difference in brain state whether the subject is on earth or twin-earth, then there is a problem for saying that the content of the experiences are different. However much they are caused by different objects, the causal relation results in the same effect.

---

<sup>30</sup> See Dretske 1995, Chapter 5.

Consider the case of a tree moved from earth to twin-earth at time  $T_1$ . From the time it reached twin-earth, the growth rings might be said to represent availability of twater. Someone studying the rings some years later on twin-earth, would not be able to tell at which point the tree arrived. Thus, there would be two indiscernible physical states, representing distinct states – availability of water prior to  $T_1$ , and availability of twater thereafter. Analogously, when Ed thinks he sees water on earth, he has a visual representation that is caused by, and co-varies with,  $H_2O$ . When he is unwittingly transported to twin-earth, Ed has an identical representation caused by, and co-varying with XYZ. Thus, Ed's physical states are indiscernible, whilst the representation is quite distinct.

This seems like a vindication of Representationalism, but there is a reason for thinking that the discussion has gone too quickly. To return to the tree: why are we justified in saying that the rings represent either 'water' or 'twater', rather than some property that these two substances have in common? According to the causal co-variation story, the most we are entitled to say about the rings, is that they represent the availability of a substance that is required for tissue growth. At the level of causal co-variation, the rings represent the same thing – presence of this property common to both water and twater; that is, on this proposal of what the rings represent, there is no relevant distinction between the tree ring types (any more than if one tree were growing in Scandinavia, and one in Peru.). What the rings represent, is cell growth, and by extension, any properties that contribute to that process. In this case, being composed of Hydrogen and Oxygen or of XYZ, is of no consequence.

Analogously, why classify Ed's thoughts as water thoughts and twater thoughts, when they could be classified according to the properties that water and twater share, that cause those sorts of brain states? Like the tree, there is a case to be made for saying that Ed's brain state is caused by some property shared by water and twater. If so, then the suggestion that Ed entertains a distinct thought on earth and twin-earth, is contradicted. Like the tree ring, his brain state represents, because causally co-varies with, some property shared by water and twater; it does not represent  $H_2O$  on earth and XYZ on twin-earth.

This leaves Representationalism in something of a dilemma. It can either embrace Externalism, and abandon its commitment to mental representations being causal co-variations with physical states in the subject's environment. Or it can keep this principle, but abandon Externalism. Neither choice seems satisfactory.

### Section Three

If we generously assume that all the above problems can be resolved, there remains the problem of the possibility of inverted qualia. Accordingly, this Section will consider how adopting a representational theory may have contributed to dealing successfully with the problem, only for it to surface once more under a different

guise. First, there will be an assessment of the arguments that purport to defuse the threat of inverted colour qualia. Second, there will be an examination of an argument that deals with all qualia inversion. Finally, the implications of this argument will be shown to generate a new challenge for Representationalism.

Subsection 1 : How intelligible is the possibility of inverted qualia?

There are at least two promising responses to Block's criticism discussed above<sup>31</sup>. William Lycan, in "Consciousness and Experience", argues that any description of the possibility is ultimately incoherent; hence, there is no problem for the Representationalist to address. Further, Gilbert Harman, in "The Intrinsic Quality of Experience" argues that such inversions would be detectable once the entire functionality of the brain is taken into consideration; hence, Representationalism has the resources to deal with qualia inversion. Both philosophers explicitly embrace an Externalist position with respect to the character of mental states, including those associated with perception. Much of what is pertinent to the debate, consists of scientific hypotheses concerning the nature of colour. As a result, I am not in a position to adjudicate upon the cogency of these hypotheses, so care is required to ensure that any particular philosophic conclusion does not rest upon the truth of any particular scientific theory.

It will be helpful to distinguish between two uses of the term 'colour'; as properties of physical objects in the world, referred to as 'physical colour'; and as a type of qualitative state, referred to as 'phenomenal colour'. The intuition lying behind the inverted qualia argument, is that there need not be any nomological link between these two types of colour. That is, although two people may be presented with the same physical colour type, one cannot infer from that fact that they will be experiencing the same phenomenal colour. Lycan subscribes to a weak form of colour realism: he assumes that physical colours do not exist as a natural kind. Instead, they may only be grouped together as properties of physical objects that produce certain sensations in creatures. So, the only physical properties that all physically red things have in common, is that they produce states of phenomenal redness in humans. As will become evident, this assumption plays a large part in his arguments.

In Lycan's opinion:

A physical colour...is a woefully disjunctive microstructural property of objects: scientifically ill-behaved...it is of interest only because of its relation to the human visual system.<sup>32</sup>

Representationalism bolsters the Functionalist account, by arguing that all experiences have intentional content that represents

---

<sup>31</sup> Chapter Three, P 48.

<sup>32</sup> Lycan. 1996. P 73.

properties of stimulations of the senses. Our experiences co-vary with the properties of the objects causing those experiences. The character of experience, therefore, is in some sense determined by the properties of those objects. The force of the qualia inversion problem, depends upon the possibility of the following situation: two individuals, both with normal brain processes, regard an object which is physically blue in normal viewing conditions; to one individual, the object will appear phenomenally blue; to the other, it will appear phenomenally yellow. Lycan argues that such a situation is unintelligible.

If such an inversion were conceivable, then it would be possible that a population may be equally divided between those who see the physical colour one way, and those who see it another. But this situation is incompatible with the notion of seeing a physical colour *correctly*; it would be arbitrary to say which half of the population enjoyed the definitive experience of physical red, for instance. Therefore, there would be no sense of 'normal' viewing conditions, or 'normal' brain processes, if these imply that physical colour is typically associated with one particular type of experience. So the suggestion is incoherent. In order to articulate the possibility of inversion, we need a stable notion of 'normal'. Yet, it is a consequence of inversion, that the term 'normal' would have no application.

However, one response to this, that Lycan anticipates, is to assert that there are intrinsic qualities of experience that are independent of their intentional content. Lycan refers to these qualities as 'Strange Qualia', and deals with them as follows. If there are Strange qualia, then they will be classified according to the physical colours to which they correspond. An individual's description of their experience will still refer to colours as such. Such a method of classification allows the possibility of there being correctness conditions of association with physical colours. However, the existence of these correctness conditions, contradicts the suggestion that Strange Qualia are independent of intentional content. So, once more, there appears not to be a coherent expression of the inverted qualia argument. Lycan wonders whether, in fact, to think about Strange qualia is still to be thinking of qualia:

The alleged Strange Qualia are collectively a set of properties related to each other as colors are, but now their names are deprived of anything like their ordinary natural-language meanings. Indeed, it is hard to see wherein they are *color* properties at all.<sup>33</sup>

Lycan considers one further attempt to set up the problem, which uses what he describes as 'New Strange Qualia'. This arises from the denial that the intrinsic qualities are classified according to physical colours. Instead, there is a phenomenal experience tied to physical properties, but somehow, there is a qualia associated with these effects. That is, they are a higher-order qualia; they are

---

<sup>33</sup> Lycan. 1996. P 111.



the 'what it is like' to have a colour experience, which cannot be expressed in terms of natural language. The problem is how to articulate the inversion problem, without utilising natural colour expressions, including reference to phenomenal colour. Lycan observes that what is so persuasive about the possibility of inversion, depends upon there being a way of describing the situation in which it is clear which qualities are being inverted, and in what respect they are being inverted. Once the qualia become ineffable, it is impossible to determine what would constitute such an inversion.

Gilbert Harman uses the idea of Representation to provide a more general argument against the possibility of inverted qualia - not just colour inversions. He argues that a number of problems concerning qualia are sustained through a failure to distinguish properties of a represented object, from those properties involved in the representation of that object. This is the same sort of error as may be made with respect to the statue of Provost Black on Princes Street, for instance. It may be said to represent a particular statesman, in which case it would have all the anatomical properties of the subject; but it would also have the properties appropriate to the material from which it was sculpted, or cast. Harman's point is that unless these two sets of properties are treated separately, you face the danger of such absurdities as saying of Provost Black, that he conducted heat inefficiently, or was liable to crack under extreme temperatures.

Harman identifies the 'argument from illusion' as being the motive for conflating the two. The argument runs as follows. Where a person sees an object X independently of X's being in the world, such as in an hallucination, it must be in virtue of X existing before the mind's eye in some sense. So, perception of the external world is always indirect. It is mediated by a direct awareness of something before the mind, called a 'sense datum'. (This argument, as Dretske noted, is responsible in part for the 'Internalist Intuition'.) However, the above reasoning is fallacious. This is demonstrated by applying this form of reasoning to other cases of non-existent objects to which we have mental attitudes, such as 'hoping for X', 'expecting that X', 'looking for X' for example. Where we are searching for some object X that does not exist, by analogy with perception, we would have to say that X has an existence in the mind. However, the person on a quest for the 'North West Passage', does not think that they are looking for something mental. Rather they are looking for something which, as it happens, does not exist to be found.

When the distinction between 'vehicle' and 'content' of representation is in place, it removes the tendency to think that by attending to the properties of an experience one is thereby attending to the representation of its subject.

So, to take the example of someone looking at a tree:

...in the case of her visual experience of a tree, I want to say that she is not aware of, as it were, the mental paint by virtue of which her experience is an experience of seeing a tree. She is aware only of the intentional or relational features of her experience, not of its intrinsic nonintentional features.<sup>34</sup>

Keeping the above distinction in mind, Harman addresses the Inverted qualia problem. He notes that our perceptions are crucial in forming beliefs about our environment and guiding our actions. So the content of a representation may be thought of as having two components, in accordance with the distinction made above. The content may be functionally defined in terms of these components: that is, defined in terms of the way that the representation is formed by the perceptual system, and also by the way the representation is used to inform our behaviour. The possibility of inverted qualia stipulates the three following conditions. First, that two people must have similarly functioning visual systems. Second, that their qualia are inverted with respect to each other. Third, that this difference is not reflected in their behaviour. Harman argues that these three conditions are not coherent, and therefore neither is the possibility of inversion.

The individuals whose qualia states are said to be inverted with respect to one another, are assumed to be functionally alike. However, there are two parts to the normal functioning of perceptual systems. One part of the system deals with the production of a representation from the sense organs. The other part deals with how perceptions give rise to beliefs about the outside world, and how these beliefs interact with other mental states, such as desires, hopes, fears and so forth. So far as the individuals are concerned, they would be using colour language in accordance with the rest of their linguistic community; they both describe blood as 'red', even though one sees it as 'green'. However, the individual who saw Red as red, would have a system which associated red experiences with Red beliefs; whereas the individual who saw Red as green, would have a system which associated green experiences with Red beliefs; that is to say, a difference between the two individuals would emerge with respect to these associative systems. Therefore it is false that the brain states of the two could be identical, and so it is false to suggest that qualia make no difference to how a system functions. The temptation to assume that the function is indifferent to the qualia involved, is enhanced by ignoring the action of the representing systems upon the belief-forming systems.

## Subsection 2 : Block - the intrinsic nature of qualia.

Even if this is an acceptable response to the inverted qualia argument, it leads to a related concern. Harman's argument relies upon the suggestion that there is nothing to an experience over and above its intentional content; that is, each experience is constituted

---

<sup>34</sup> Harman. 1990. P 39.

by the properties of the object it is representing. There are no intrinsic non-intentional properties. It is this suggestion that is attacked by Block in a series of articles, beginning with "Inverted Earth"<sup>35</sup> and continuing to the present. Block wants to preserve a distinction between, on the one hand, the intentional or representational content of an experience, and on the other, its qualitative or sensual character. For instance, when looking upwards into a cloudless sky, one has an experience of 'blue'. Block's claim, denied by Representationalists, is that not only does the experience bear information about the object of awareness, but there is also a quality of that experience - its *quale* - which is in some sense independent of its intentional content. If Block can prove this, then it undermines Harman's argument against inverted qualia, as well as overturning a fundamental premise of Representationalism.

In "Mental Ink and Mental Latex", Block suggests that there are three aspects to 'experience': that there is an intentional content, whereby objects are represented; there is a mental property that does this representing - mental ink; and that there are mental properties that represent nothing at all - mental latex. His attack upon Representationalism is based upon this division. Success depends upon his being able to provide convincing examples of mental ink and latex. This he attempts by demonstrating that it is possible to have a change in content that is not accompanied by a change in character, and by providing instances of experiences that lack representational content. The following is a brief examination of these attempts.

One way of showing mental ink exists, is by adapting the twin-earth thought experiment originated by Putnam. Block takes the idea of this experiment, and supplements it by having twin-earth as a colour inversion of earth. Hence, where we would describe a particular object as being yellow, on twin-earth that same object would appear to us as if it were blue. However, the inhabitants of twin-earth employ a different colour vocabulary from earthlings; for each colour term we would use to describe the hue of an object, they would use the term that corresponds to the opposite colour. Thus, where we would describe the colour code for the earth wire as being yellow and green, they would say that it was blue and red. The result of these conditions is that on earth, a typical speaker of English would have a particular type of visual sensation whilst gazing out over the Sahara desert, which they would call 'yellow'. On twin-earth, a typical speaker of twinglish looking over the twin Sahara, would also have a sensation that they would call 'yellow'. Nevertheless, if the individuals changed places, they would have different sensations, which they would both describe as being 'blue'.

Imagine that the person from earth, Ed, wears lenses that invert the rays as they enter the eye. Whilst wearing these, all colours are experienced as their opposites. He is then sent to twin-earth. When he visits the desert, he has an experience that he naturally describes as 'yellow', in conformity with the twin-earth

---

<sup>35</sup> Block. 1990.

inhabitants; even though the sand would normally look to him as if it were 'blue', his inverting lenses have altered the wavelength of the light so that they now correspond to that of yellow. After a while of living in the community, Ed forms the same sort of beliefs as everyone else on twin earth; that is, he uses colour terms in conformity with the other inhabitants. Functionally Ed is indistinguishable from other people, though the character of his experiences is quite distinct; that is, those conditions that give him what he would describe as a 'yellow' experience, are giving to the others an experience that he would describe as 'blue'. That is, as Ed uses the term 'yellow', there has been a change of intentional content without there having been a change of character. Therefore the intentional content of an experience cannot be identified with its character.

It is unlikely that either Lycan or Harman will accept the above as proof of there being 'mental ink'. By placing inverting lenses over his eyes whilst on twin earth, Ed can hardly be said to be using colour terms in the same way as the rest of the community. Generally, in order to allow a comparison between the appearances, the objects and the viewers have to be specified as being in standard conditions of lighting and physical well-being. This thought experiment ignores these stipulations completely, by giving Ed non-standard viewing conditions. Ed is regarding the object through inverting lenses; his twin-earth counter-parts have no such mediation between the desert and their retinas. If he were truly to be using colour terms in accordance with the twin-earth community, then he would have to describe his experience as "yellow as seen through inverting lenses". This usage is not equivalent to straightforward "yellow", which is how the community would describe the colour of the desert.

Block's attempts to produce instances of 'mental latex' are equally unconvincing. Although such phenomena as orgasms and phosphene experiences imply little in the way of representative content, it is not to say that they represent nothing at all. These instances illustrate a truism about representation. If you grant that sensory experience is the product of physical stimulation, then trivially, each experience will carry, as a minimum, the information, that such and such a process is being activated. There is possibly a confusion between 'bearing information' on the one hand, and 'being informative' on the other. Whilst the latter implies the former, the converse is by no means true; a sensation in my back can convey the information that my latissimus dorsii muscle is being over-stretched. Having no knowledge of anatomy or physiology, it is just a pain in the back to me. This confusion applies to Block's putative examples of 'mental latex'. Simply because the orgasm or phosphene experiences signify nothing to the experiencer as such, does not mean that it is impossible to derive knowledge from them. One suspects that the vocabulary of orgasms is attenuated due to cultural delicacy, rather than their being without representational content.

### Subsection 3 : Has the riddle of inversion been solved?

However, the position taken by Harman and Lycan may be vulnerable from a different direction. They assume, like all Externalists, that the character of a sensation is determined by its external cause. As seems highly plausible, the character of an experience is determined by the relation between the physical properties of the stimulus upon the physical properties of the creature's sensing mechanism. If two people of identical physical constitution, are placed in identical environments, the above assumption allows no room for there to be any difference in their experiences. For if the experience is a physical event, then any difference of experience would have to be explained in terms of a difference in their constitution. By hypothesis, there is a sameness of constitution, so the possibility of spectrum inversion only makes sense if one denies the principle that like properties have like effects. That is, even though their physical descriptions are relevantly the same, nonetheless their qualia may be different. Clearly this suggestion only has force against Materialists, like Block, who are antipathetic to Externalism.

The following consideration suggests that the inversion problem has not yet gone away. Harman's point against Block is that when the total state of the brain is taken into consideration, differences will emerge by which one would be able to detect an inversion. Consider again the example of the forming of the belief that 'strawberries are red'. The idea is that there will be distinct processes involving red and green, and that these are linked to other processes in the formation of beliefs. In order for an inversion to be possible, there will be green processing associated with the formation of red beliefs, as opposed to the completely different red processing being associated with the belief mechanism. This difference would emerge, and so would not count as a proper instance of inversion.

There is a crucial ambiguity between 'green sensing' and 'green processing'. The latter refers to the scientifically discoverable activities of the brain, which correspond to the presentation to a normal subject of a green object (where green is defined in terms of certain physical constraints). Thus the composition of the brain is classified according to features from a third person point of view, rather than the agent's experience. My belief that it is 'green' that I am seeing, is based upon the same linguistic usage that decreed that 'green' processing takes place in such and such a region of the brain: but that usage has been determined independently of the quality of my experience. It appears that 'green processing tissue' is so-called because of its association with the properties of certain external objects - call them 'G' type properties. To assume that the quality of experience will be the same, is thus ultimately based upon the fact that the subject is presented with an object with 'G' type properties. This arbitrarily disallows the possibility that the tissue actually produces red

experiences. In short, it simply begs the question against the possibility of inversion.

This point is a corollary of Kirk's point, that the tissue in the brain will produce experiences when stimulated, but one cannot assume that the same distal stimuli will produce the same effect. Consider the physical nature of hair; if one were investigating its structure and lived amongst, and only knew of black-haired people, one would be excused for overlooking the possibility that the same basic structure could produce hair of different colours. Because one assumed that all hair was black, so one would be blind to the properties of hair cells which determine the colour of hair. Similarly, the differences in tissue which might be responsible for producing different types of experience, are concealed by the fact that one assumes that sameness of external property implies sameness of experience.

There is a point to make against Lycan's discussion of qualia. The conclusion that the properties to which strange qualia refer are ineffable, is not itself a proof that strange qualia do not exist. Certainly, one must accept that with the complete breakdown in colour terms, reference to these experiences would be impossible in language: but that may be a consequence of language generally, rather than one of existence. In a way, there is something quite poetic about the first-person acquaintance with experience being ineffable; indeed it reflects the idea that we never will be able to share the quality of experience with anyone else. Yet I do not see this as implying that no-one else's experiences exist. The incoherence of the expression of the possibility of qualia inversion, is similar to the incoherence of the expression of the sceptical position. This incoherence though, is compatible with the truth of these positions.

#### Section Four

The final difficulty for Representationalism that requires consideration, is the problem of 'Misrepresentation'. This problem concerns the discrepancies between my perceptual experiences, and my beliefs that result from those experiences. Mental states are said to represent in virtue of their causally co-varying with the objects of which they are representations. Yet consider the following common example. Whenever I raise my hand toward my face, it appears that my hand grows in size. Similarly, this occurs when other hands approach my face. However, whenever hands approach other people's faces, they either stay the same, or in the case of my hand, they diminish in size. Even though my experience represents my hand as changing shape, I have the belief that, anatomically, the position of any hand relative to my face has no bearing upon the physical mass of the hand. Hence these representations are at odds with each other; my perceptual state represents something that changes shape, whilst my belief state concerning that object, represents something that stays the same.

If our conscious states co-vary with the physical properties of our environment, then there seems to be no room for the possibility that our perceptions misrepresent the world: phenomenal experience is simply a function of the relation between the environment and our sensory systems. In this respect, our ability to get things wrong becomes quite curious. As Dretske says:

How is it possible for physical systems to *misrepresent* the state of their surroundings?<sup>36</sup>

Dretske discusses the problem of how certain states of mind get their meaning. A sensation of pain represents tissue disturbance, for instance – that is what the sensation means. The problem is, that we assume that beliefs can be false. Representationalists regard the contents of beliefs as representations. Therefore, if the belief is to be false, then the content must be a misrepresentation. Yet, how such a circumstance could arise in a natural relation of causal co-variation, is a mystery. Somehow, there has to be an account of how this relation can go wrong, if there is to be any good reason to accept Representationalism.

Consider the example of the magnetosome in certain marine bacteria, that can only live in oxygen-free waters. Magnetosomes allow the bacterium to move toward a magnetic source, by producing a certain stimulus – call it M. In their natural environment, behaviour stimulated by M has the consequence of allowing the bacterium to navigate toward regions of the ocean where there is a lack of oxygen. M could be regarded as having the functional meaning D: “direction of oxygen-free water”. However, the bacteria are also sensitive to other sources of magnetic attraction, which need not be associated with oxygen levels. If a magnet is held before the bacterium, thereby producing M, can we then say that D misrepresents the direction of a safe environment? Dretske observes that this interpretation will not suffice to provide a natural account of misrepresentation. The fault lies with the particular attribution of the meaning D to M. There is no justification for choosing this particular meaning instead of, for instance, D': “direction of magnetic source”. Clearly, D' is not a misrepresentation of the state of affairs.

Dretske's solution to this problem is to consider a creature of greater complexity than the bacterium. There are two processes by which this organism responds to the presence of a particular object O in its environment. Accordingly, there are two stimuli representing the presence of O – call them S<sub>1</sub> and S<sub>2</sub>. If O is a predator, then S<sub>1</sub> and S<sub>2</sub> can be regarded as meaning ‘predator’. These stimuli may become linked to an avoidance behaviour state, called B. As a result of this linkage, B also means ‘predator’, granted that B occurs whenever either of S<sub>1</sub> and S<sub>2</sub> occur.

Suppose further that the organism is capable of associative learning; that is, to associate a conditioned stimulus (CS for short) with a particular object or event. In this situation, B is triggered by

---

<sup>36</sup> Dretske. 1994. P 157.

some CS. Since any stimulus can potentially become a CS, then the meaning of B cannot be tied to any particular stimulus. However, it will still have the natural meaning of 'predator', since that has been conferred upon it by the associative processes. Now it is possible to see how a natural meaning can misrepresent. B represents O whenever it is triggered by a stimulus that has been associated with Os, but which does not invariably accompany Os. On the occasions when that stimulus has not been prompted by an O, it triggers B, which is now a misrepresentation of O.

There are two difficulties with this account. Firstly, can the presence of two stimuli make the difference between an organism in which misrepresentation is not a possibility, and an organism where it is a possibility? Secondly, by utilising the processes of associative learning, is this solution available to Representationalists?

The first difficulty arises as follows.  $S_1$  and  $S_2$  are taken as representing aspects  $A_1$  and  $A_2$  of O, though  $A_1$  and  $A_2$  are not the same. Either,  $A_1$  and  $A_2$  are exclusive to O, or there may be some other entities that instantiate these properties. In the first case, the stimuli will always indicate the presence of O; that is, there will be no occasion where these stimuli misrepresent the presence of O, since they are triggered only by the aspects  $A_1$  and  $A_2$ , that are exclusive properties of O. In the second case, if  $A_1$  and  $A_2$  are also properties of P, then the stipulation that  $S_1$  and  $S_2$  mean O, will be as unmotivated as the stipulation that magnetosome activity meant the presence of oxygen-free water. Again, the suggestion that  $S_1$  and  $S_2$  are misrepresentations of O, is unwarranted. So in either case, there does not appear to be a case of misrepresentation: and multiplying the stimuli will not avoid the above dilemma.

The second difficulty pre-empts a move that might be made against the above criticism. It concerns the status of learning in this solution. One can appreciate how misrepresentations occur. We take one thing to signify another, because we have associated the presence of the former with the latter. For example, the famous red sky at night so beloved of shepherds the world over, is taken to signify fine weather. However, it is possible that the redness is caused by a distant inferno, in which case the following day may well be overcast. On this occasion, the redness misrepresented the state of tomorrow's weather.

The failure of the association was due to an invalid inference from the colour of the sky to the weather. This inference is an extra step in the process of representation. First redness, 'A', is represented. Separately, fine weather, 'B', is represented. What makes the redness represent the weather, is the inference that  $A \Rightarrow B$ . Representationalism must be able to provide an analysis for this third belief, since it is this type of erroneous belief that is doing all the work in the above solution. Yet, how does a Representationalist account for this third belief? All mental states are representations, so a plausible candidate for this belief, is the causal co-variation of As and Bs: but why should this be taken to mean ' $A \Rightarrow B$ ', rather than 'A is sometimes associated with B'? In other words, how can the



natural sign that underpins this third belief, indicate the falseness which ultimately leads to the misrepresentation? The answer, once more, is surely, that there is no motivation for thinking that this belief has the false meaning ' $A \Rightarrow B$ '. Representationalists are obliged to give an account for how this third belief can misrepresent. This they have not done. Thus, so far, it looks as if Representationalism lacks the resources to deal with the familiar notion of misrepresentation.

Representationalism has encountered stubborn resistance in its treatment of cognition and sensation. Its problems with how beliefs qua natural signs can go wrong, the lingering possibility of inverted qualia, and the uncertainty surrounding the details of how brain tissue can realize states of sensations, combine to dismay the Materialist. It is with a sense of humility that the next chapter considers the following possibility; that human intelligence is incapable of conceiving how the body and mind are related.

## Chapter Five

Humility is all very well in itself: but it should come as no surprise that when philosophers become humble, their self-abasement is accompanied by reflection upon the grounds for taking up such a position. It is easy to comprehend why one might concede that the problems of consciousness are simply too difficult for humans to solve. There are instances in the literature which illustrate this idea. Nagel<sup>1</sup> likens us to an ancient Athenian who is told that 'matter is energy'; the equation may be faultless, but the terms it employs are so far removed from the listener's comprehension, that the statement is simply uninformative. Similarly, Jackson<sup>2</sup> compares us to creatures with limited abilities - sea slugs - who form a cosmology based upon their current understanding of the environment. One may sympathise with their attempts, no matter how inadequate they are. Yet this sympathy is misplaced; we are no nearer a satisfactory understanding of life than the slug.

At this point, any reasonable thinker might be caught by a feeling of guilt at the presumption of being able to make clear all that goes on in the world. The next step is to renounce any further philosophic effort, and remain mute concerning the mystery of it all.

This may well be an understandable reaction - maybe even reasonable. Philosophically, it simply will not do.

The purpose of this chapter is to examine the attempts to render an agnostic position philosophically worthy. In the main, it will concentrate upon the work of Colin McGinn. If successful, the arguments will allow one to acknowledge the strangeness of the mind, without thereby surrendering to a belief that it is immaterial. In that sense, it will have provided a solution to vexed questions about 'qualia', 'intentionality' and 'personhood' with which Materialists have been embarrassed; for the solution will be that the nature of the mind - albeit material - has only the illusion of immateriality. This illusion is generated by limitations in our cognitive abilities, and should be treated as such. Just as the straight stick appears bent when regarded through water, so too, the mind appears mysterious when regarded by our feeble intelligence.

Sadly, there are reasons to doubt that this is such a straightforward position to adopt. The first Section of this chapter will deal with a number of arguments in support of the idea, that there are features of the mind whose nature cannot be captured by Materialist ontologies. These examples will motivate the contention of the second Section, that there is an 'Explanatory gap' between the physical and mental. In Section Three, McGinn's ideas on 'New Mysterianism', as Flanagan calls it, will be examined in some detail, before finally concluding in Section Four, with some objections to this position.

---

<sup>1</sup> Nagel. 1974. P 447.

<sup>2</sup> Jackson. 1982. Pp 135-6.

## Section One

Is it possible to describe the mind in physical terms? Two influential writers, who doubt that such a description is available, are Thomas Nagel and Frank Jackson. Nagel is concerned that the very nature of our understanding stands in the way of our ever being able to provide such an account; he recognises that, as such, his view is compatible with the truth of Materialism. Jackson's argument challenges the idea that Materialism is able to articulate truths about the nature of experience; however, unlike Nagel, Jackson sees this failing as pointing toward an anti-Materialist position. There is much that has been said concerning the validity of these positions<sup>3</sup>, but for now, the purpose of presenting them in outline is to place McGinn's arguments in context. He is influenced by Nagel, but would deny Jackson's conclusion. The point of interest, is to see whether acceptance of Nagel's view is consistent with the denial of Jackson's, given the similarities between their positions.

### Subsection 1 : Nagel – could there be an objective account of 'subjectivity'?

Nagel argues that there is a tension between our conscious apprehension of the world, with the theoretical understanding of the world that we generate, for instance, through science. This view is developed in particular through the essays, "What Is It Like to Be a Bat?" and "Subjective and Objective", as well as the book "The View from Nowhere". A survey of this position is based upon these sources, to demonstrate why he thinks there is an insurmountable obstacle between us and a complete understanding of the world we inhabit. The challenge for the Materialist is to give an account of how any understanding of the workings of the mind is possible. If one assumes that only scientific practices provide us with any kind of understanding, then Nagel suggests that, for two areas in particular, the nature of 'Mind' will remain a mystery: these areas concern the relationship between the mind and the body, and the special relation I bear to myself.

Nagel observes that there is a particular pattern to the way we form theories about our world. At first, we gain information about the world through our senses. This provides a view of nature from a subjective standpoint; which is to say, we are acquainted only with those properties of objects to which we are sensitive. Once we begin to categorise the objects of perception, it is a natural progression to suppose that those properties that give rise to perceptual experiences in humans, may have different effects upon other organisms; further, that such properties would exist regardless of there being any entities sensitive to them. Hence, the true nature of objects is independent of their appearances, and indeed, need not bear any resemblance to those appearances. A greater understanding

---

<sup>3</sup> For two influential responses, see Lewis 1992a; Nemirow 1992; Alter 1998. For a concise summary, see Van Gulick 1993.

of those objects may be arrived at by theorising about other properties that objects might instantiate, even though these new properties are such that our senses are not activated by them. The further we move away from the subjective perspective of these objects, the more complete our understanding. For instance, our understanding of gravity has to move beyond the feeling of heaviness in our limbs and the sight of objects falling to earth, if it is to do justice to such an essential part of the way matter is composed; that is, it has to encompass the unseen nature of the forces of attraction between bodies:

We flee the subjective under the pressure of an assumption that everything must be something not to any point of view, but in itself. To grasp this by detaching more and more from our own point of view is the unreachable ideal at which the pursuit of objectivity aims.<sup>4</sup>

Our understanding of consciousness should therefore proceed according to the above method. Yet a moment's reflection is enough to recognise that 'Consciousness' is a remarkably slippery concept to articulate. One of the problems faced by Identity theorists, was precisely the difficulty of finding a criterion that was neither too Liberal nor too Chauvinist. Nagel suggests that a useful criterion for consciousness, is that where it is present in an organism, there is something it is like to be that organism. The 'what it is like' is no less than the subjective character of experience, which serves as the starting point in our theorising about the nature of what there is. Therefore, to provide an understanding of consciousness, we need to adopt an objective view of this subjective character: and that, argues Nagel, is simply unintelligible.

Firstly, that we should appreciate the difference between the subjective and objective views, he considers what it would be like to be a bat. The creature is chosen because we may be reasonably certain that it has states of consciousness; that it is not too close to humans to allow for our imagination to extrapolate from our experiences of hearing, for instance, to what their experiences of hearing might be like; and particularly because bats have a sense of echo-location which we do not possess. However much we know about the bat's perceptual abilities, in terms of structure and function, there still remains the question of what that experience would be like. No matter the extent to which we picture ourselves as using a psuedo-echolocatory device - such as a stick - there is no justification for assuming that what we then experience would be an accurate simulation of a bat's experience: no more would we accept that a fish might think it was experiencing the varied *tastes* that we enjoy, merely by marking the varieties of food according to differences in their *texture*. Though our discriminative abilities would be on a par, the discriminations would be based upon entirely different types of experience.

If the idea of being a bat is too outlandish, the distinction between what the objective and subjective views tells us can be

---

<sup>4</sup> Nagel. 1979. P 208.

illustrated by comparing what it would be like to be a patient in Bach y Rita's experiments with blind people<sup>5</sup>, whose skin is stimulated according to light patterns, and who claim to be able to 'see' rather than 'feel' objects. Clearly it is an inconclusive argument, but personally, I have no idea of what it must be like to have the familiar experience of tingling skin, transformed into something like a vision. However, the important point is that given the same objective information about a particular experience, one would expect people of equal intelligence to have an equal understanding of that experience. Yet this seems not to be the situation when we attempt to appreciate what perception must be like for the bat or the blind patient:

Reflection on what it is like to be a bat seems to lead us, therefore, to the conclusion that there are facts that do not consist in the truth of propositions expressible in a human language. We can be compelled to recognize the existence of such facts without being able to state or comprehend them.<sup>6</sup>

If the distinction between the objective and subjective views is to be maintained, then an attempt to understand the nature of our experiences is attempting the impossible: if 'understanding' is the systematic moving away from the subjective view, how ever will we form an understanding *of* that subjective view? If objectivity is achieved by moving away from appearance to reality, then our study of 'appearance' will not be furthered by this practice. The constraints of objectivity require that we distinguish between how things appear, and how things are. There is no doubt about the existence of those experiences that are the foundation of our understanding. The problem is how to articulate a description of their true nature. Science is successful because it removes the infidelities of appearance; in its purest form, it eschews all influence from the senses. Such a method, whilst in many ways salutary, has an obvious failing:

...physics is bound to leave undescribed the irreducibly subjective character of conscious mental processes, whatever may be their intimate relation to the physical operation of the brain.<sup>7</sup>

Therefore there is no intelligible possibility of providing an explanation of 'what it is like', which is to say, of 'consciousness'.

Nagel believes that the way of overcoming this problem, is to accept that the 'physical' interpretation of reality is not exhaustive; it may be complimented with an objective view of the mind. Our understanding of the mind would proceed just as our understanding of physical objects proceeds - by abstracting from our theories the personal standpoint that we occupy. Thus, in the first place, mental concepts would be articulated as they might be experienced by any other human; then, as experienced by any other organism; and so on,

---

<sup>5</sup> As described in Dennett 1991. Pp 339-342.

<sup>6</sup> Nagel. 1974. P 441.

<sup>7</sup> Nagel. 1986. P 7.

to a point where these concepts reach a level of generality that they apply to every type of organism capable of consciousness. Just as executing this procedure for generating an understanding about the physical world failed to account for the facts of subjective experience, so too, does this procedure have its deficiencies. In order to understand any fact about the world, an individual has to adopt as objective a viewpoint as possible. There are facts concerning the nature of every individual whoever existed. What seems to be unaccountable, are those facts which identify me with the particular individual I am. Having scrupulously stripped away all features of my experience which would contaminate any theories that I could form about the world, there is no longer any provision for associating myself with any of the objectively described individuals in the world.

There are two aspects to this problem: firstly, that the objective view is incomplete with respect to the facts that require explanation; secondly, that there are no criteria for identifying the subject of experience with any particular individual; Nagel remarks that the proposition "I am A.R", normally taken to express an identity relation, has now more the look of a subject-predicate expression. This arbitrariness is a consequence of treating experience as neutrally as possible; that is, as thinking of the world from no particular point of view at all. This way of thinking, however, ignores the important fact that my consciousness occupies a particular standpoint. For those attempting a complete understanding of the world, the following dilemma has to be confronted. Either accept that objectivity is a genuine method to discover truth, and live with the following unsatisfactory consequence: that this method is blind to facts about the nature of our experiences and of ourselves as conscious beings. Or deny that objectivity is the proper method, and hope to discover a replacement. This alternative method will have to be faithful to our intuitions about how to achieve knowledge - the very intuitions which selected the process of objectification - and explain all the putative facts of the mind. Neither option affords grounds for expecting an easy solution to the mind-body problem.

### Subsection 2 : Jackson and 'the knowledge argument'.

The other example from the literature that has had a great influence upon this matter, comes from Frank Jackson. In a pair of articles, "Epiphenomenal Qualia" and "What Mary didn't Know", Jackson employs what he calls 'the knowledge argument' to show that there are:

...certain features of the bodily sensations especially, but also of certain perceptual experiences, which no amount of purely physical information includes.<sup>8</sup>

---

<sup>8</sup> Jackson. 1982. P 127.

This conclusion is supported by the outcome of two thought experiments - one slightly fantastic, but the other more credible. The first involves a character called Fred, who is sensitive to more colours than we are; in cases where we would assert that the colour of a brake light and blood were the same, Fred would say that actually we were mistaken, and that the two shades were quite distinct (referred to as  $red_1$  and  $red_2$ ). In effect, Fred's abilities relative to ours, are analogous to ours with respect to colour-blind individuals. Importantly, when investigated, scientists are able to identify structures in Fred's visual system which could account for his extra discriminatory powers, just as it is possible to give a similar account for colour-blindness. To that extent, there is nothing mysterious about Fred physically; however, the force of this example is to demonstrate that no matter how complete the physical information concerning our visual processing, we would still want to know what the colours  $red_1$  and  $red_2$  look like. Therefore, Jackson concludes, facts about the nature of experiences, cannot be given in physical terms.

The other example requires less imagination, and for that reason perhaps, is the example which receives most attention. Mary is a scientist studying the nature of colour in all its manifestations. By hypothesis, she knows everything there is to know about the physical nature of colour. However, her life has been spent in a monochrome environment, such that every object with which she is familiar by sight, is either white or black, or a shade of grey. The question Jackson poses, concerns Mary's experience when she is allowed to leave her laboratory, and acquaint herself with polychromatic objects. If Materialism is correct, then she should know everything there is to know about colour, including what it is like to see a lemon as yellow, for instance. Yet, just as we would dismiss the idea that a congenitally blind person could appreciate the experience of vision by research, so too are we inclined to acknowledge that Mary would be ignorant of the experience of seeing something yellow, until she emerged from the laboratory. Once again, the conclusion is that even when in full possession of the physical facts, only first-person acquaintance with colour can provide knowledge of what it is like.

For Jackson, these arguments are sufficient to persuade him of the falsity of Materialism, and the truth of Epiphenomenalism. It is a pertinent consequence of his position that he sees Materialism as being too optimistic in its ambitions, with regard understanding the world and ourselves. He presents the following analysis of why the nature of qualia is so poorly understood. Humans are a product of Evolution; features of our nature may be explained in terms of how they contribute to our ability to survive. Since qualia are Epiphenomenal, they play no part in our survival - they have no contribution to make simpliciter. Therefore, any faculties that would increase our ability to understand qualia, would not be selected, and would therefore not appear in evolved species. Granted how little we need to know in order to survive, relative to the amount there is to know, it is not surprising that there are so many gaps in our

knowledge. Since such knowledge is irrelevant to our species' continuing existence, there seems little chance that we will ever come to arrive at answers to these questions. Therefore, not only do qualia elude the Materialist's fumbling attempts at comprehension, but the very constraints of our existence make it highly improbable that we will ever come to understand their real nature.

The relevance of the above to McGinn's position is that both he and Jackson accept that this understanding is beyond us, and both accept that this is due in part to pressures of survival; but whereas McGinn believes that the mind has a physical nature, Jackson argues that it is a non-physical entity, that wholly escapes our cognitive abilities. It is essential to the New Mysterian position that it establishes how the sense of mystery arises naturally out of our understanding of physical properties.

## Section Two

What emerges from the above considerations, is the need to establish the connection between how a property like consciousness is composed, and how we get to learn about the nature of this composition. Jackson's argument proceeds from a premise concerning the impossibility of an account of qualia in physical terms, to a conclusion that qualia are not physical properties. Whilst his thought experiments provide cogent support for such a conclusion, it is far from clear how the claim that 'physical facts are insufficient for understanding the nature of qualia', warrants the assertion that 'qualia are not physical'. The Materialist must commence their defence by showing that Jackson has no metaphysical conclusion to draw, and take their chances with the epistemic consequences of the Mary case.

This defence is clearly presented by Joseph Levine in a series of articles<sup>9</sup>, where he introduces the idea of the 'Explanatory Gap'. He notes that Jackson's argument focuses upon the following claim. For any putative physical description 'P' of a mental event 'M', it is conceivable that 'P' could be accompanied by a different mental event, or no mental event at all. In this respect, it belongs to a family of arguments that demonstrate that for any property of the mind 'M', whichever physical property, or functional process 'P', you wish to identify it with, it is always conceivable that 'M' is not identical with 'P'; or, in the extreme case, that 'M' is not identifiable with any physical property, or process, at all.

At the most general level, the argument can be made thus. It seems coherent to suppose that I might not have a physical being - I might be in the grip of an illusion. However, it does not seem coherent to imagine that I do not have a mind, since the existence of a 'mind', is presupposed by any act of imagination. In this sense, the mental and the physical are distinct, and therefore the mind is not physical. Levine deals with two contemporary versions of this form of argument, which use the criterion of conceivability to establish an anti-Materialist position; these are the arguments from

---

<sup>9</sup> Levine. 1983, 1993 and 1998.



Kripke<sup>10</sup> and Jackson<sup>11</sup>. Neither establish the metaphysical conclusion they seek, but for all that, Levine argues that the epistemic conclusion that can be drawn is just as bothersome for a Materialist.

Subsection 1 : Levine's remarks on conceivability arguments.

Without going into the details of Kripke's argument, its conclusion is as follows. If an object referred to by one description is identical to an object referred to by a different description, then they are identical of necessity. For instance, "the sum of seven and twelve", picks out the same number as "the prime number between seventeen and twenty-three". This implies that in every possible world these descriptions pick out the same object. However, there are some putative identities where it appears one can imagine worlds where the relation fails: that is, where the descriptions pick out different objects. Unless one can give an account for why such a failure is only imaginary, rather than actual, one has to conclude that the identity was entirely spurious. One such set of putative identities, is that which lies between the mental and the physical. Kripke suggests that for any identity relation between these that we care to suggest, it is always coherent to imagine its not holding. Therefore the mental and the physical are not identical.

To make these remarks clearer, consider two superficially contingent identities:

"Water is H<sub>2</sub>O"

"Pain is the firing of C-fibres".

There are two important differences between these statements. The first being that when we imagine water not being H<sub>2</sub>O, what we are really doing is imagining something that resembles water not being H<sub>2</sub>O. However, when we imagine pain being something other than the firing of C-fibres, such a move is not available. We cannot be imagining something resembling pain, without that something being painful. In the water example, we can distinguish between its appearance and its true nature; whereas in the case of pain, its appearance seems to be its true nature - nothing could feel like pain and not thereby be painful. So contingency in the first statement is explicable through there being a distinction between appearance and reality, whereas there is no such distinction to be made in the second statement.

A second difference lies between the above statements, insofar as it is possible to imagine both that you could have C-fibres firing that did not involve pain, as well as having a pain that was not associated with C-fibre activity. However, whilst it might be possible to imagine water not being H<sub>2</sub>O, it does not seem possible that H<sub>2</sub>O could not be water - where 'water' picks out a set of appearance properties. Levine suggests a reason for this. Our knowledge of physics and chemistry constrains how we can intelligibly imagine certain properties manifesting themselves; given the properties of Hydrogen and Oxygen, it is inevitable that

---

<sup>10</sup> Kripke. 1980.

<sup>11</sup> Jackson. 1986.

they behave the way they do. On the other hand, it is conceivable that our sensitivity to these properties was different, so that water would have a different appearance to our senses. In the case of pain, there is no knowledge of the properties of neuro-physiology that provide the same level of intelligibility of the link between brain tissue and pain experiences:

...and what is left unexplained by the discovery of C-fiber firing is *why pain should feel the way it does!* For there seems to be nothing about C-fiber firing which makes it naturally “fit” the phenomenal properties of pain, anymore than it would fit some other set of phenomenal properties.<sup>12</sup>

The chances of coming across some suitable physical candidate for this position, strikes some as non-existent. As Jackson’s ‘knowledge argument’ demonstrates, completeness of physical information is compatible with ignorance of the nature of psychological states.

Much has been written criticising these views of Kripke and Jackson; yet even if we concede that they are correct in their remarks, Levine holds that they have mistaken an epistemological point for a metaphysical one. So far, all that has been demonstrated, is that there are certain facts about the relation between the mental and the physical, to which we have no access. Our ability to imagine states of affairs, of itself, fails to imply the anti-Materialist conclusion that the mind is not essentially physical. All it shows, is that there is no means of understanding why the physical underpins the nature of the mental, in the way that we understand how the nature of water is determined by its physical constitution. In order to thwart the Materialist, there has to be an additional argument to show that all metaphysically possible states of affairs are epistemically accessible; if this were so, and it were demonstrated that there was no way of knowing how the mind was physical, then the metaphysical possibility of Materialism would be denied. However, the initial premise of this argument is not forthcoming.

There is at least one other approach that an anti-Materialist might attempt. The deflection of the metaphysical argument, depends upon there being distinct paths of access to the same fact. All the above considerations show is that the first and third person paths may be distinct but need not be thereby providing knowledge about distinct objects. However, there may be the seeds of a problem here for the Materialist, depending upon how one analyses this notion of epistemic access<sup>13</sup>. Much depends upon the following claim. In cases of identity, what is identified is a single object known by two different paths; that is, two distinct properties are picked out independently of our knowledge of their being properties of the same object. These properties have to be contingently related. If they were related in virtue of the meanings of the descriptions by which we refer to these properties, then it is nonsense to suppose that the identity would be discovered empirically, rather than by

---

<sup>12</sup> Levine. 1983. P 357.

<sup>13</sup> Referred to as the ‘Distinct Property Model’ or ‘DPM’. See Levine 1998. P 455.

analysing the meanings of the terms. In addition, any stipulation of a non-contingent relation between these sets of properties, would seem to beg the question against the conceivability argument, as described above.

For instance, consider the identity between Hesperus and Phosphorous. These appearances of Venus have distinct spatio-temporal properties by which they were originally singled out. These are contingent properties of Venus: the relative orbits of the Earth and Venus around the Sun, make the latter planet appear at different times of the day. Similarly, it is hoped that the identification between 'pain' and C-fibre activity can avoid the implications of the conceivability argument, by using DPM. Levine doubts that this approach will be successful. There are at least three related reasons to suppose that the Materialist be unable to resort to this strategy.

First, if we are to make a similar identification between 'functional state of the brain' and 'qualitative state of the mind', then we have to accept that there are two distinct ways of discerning these states; and indeed, we may be happy to associate respectively third person and first person access with these states. However, if there is a correspondence between type of state, with type of access to that state, then we are forced to assume that there is an object with distinct functional and qualitative states. This conclusion would not suit the Materialist, since it follows that whatever we know about the functional properties of the mind, such knowledge would not inform our understanding of its qualitative properties.

Second, unless they are Epiphenomenalists, Materialists are committed to the idea that qualia have a rôle to play in behaviour and the formation of beliefs. The way we sense the world influences what we believe to be the case, and how best to act in order to satisfy our desires. Therefore, the character of experience has in some sense an explanatory rôle with regards what we believe, and how we act. Thus Materialists cannot accept the above conclusion that functional states are at best only contingently related to qualitative character; to do so would sever the explanatory link between feeling and action. It is the frightening quality of 'pain' that explains our behaviour whenever we are hurt, as well as explaining why we take steps to avoid pain. 'Pain' is not contingently frightening - it could not have the quality of being soothing, for instance.

Third, the employment of DPM either leads to an infinite regress, or to an acceptance of irreducibly mental properties. This may be shown as follows. There is a single object 'O' whose properties may be sorted into at least two sets, 'A' and 'B'. 'A' and 'B' share a small number of properties; small enough to make it reasonable to suppose that someone could know enough about 'A' and 'B', without immediately realising that 'A' and 'B' were aspects of 'O'. As knowledge increases, so too does the chance of discovering these shared properties, and hence making the identification. 'A' and 'B' refer to the same object, even though they do it by distinct representations of that object. So, a

Materialist can hold that 'pain' and 'C-fibre activity' refer to the same state, because they are distinct representations of that state. However, in order for psycho-physical identification to be made, it has to be possible to discover what 'painfulness' and 'C-fibre activity' have in common, to justify believing that they are both aspects of a single state

This empirical discovery will inevitably involve the identification of mental and physical properties; that is, in order to find out what the above 'have in common', will require a comparison of their properties. In order for DPM to work, there should be some physical properties of C-fibres that are the same as properties of painfulness: call them 'C' and 'D' respectively. Yet every putative identity between 'C' and 'D' will be susceptible to the conceivability argument. One can appeal to the DPM strategy again to explain why 'C' and 'D' are the same although they seem different, but then one must come up with properties that 'C' and 'D' have in common. Granted the acknowledged difference in mental and physical properties, there appears to be no end to the number of times DPM would be employed. Either there is an infinite regress, or one must accept that there are irreducibly mental properties.

Levine suggests that there is a way of re-stating the Distinct Properties Model that avoids the above implications. He notes that there are two ways in which the world presents itself to us; as an active process, whereby we attach concepts to the contents of our experience; or as a passive process, whereby properties of our environment impinge upon our senses to produce experiences. With the first way, which he calls the 'ascriptive mode', reference is fixed according to the extent to which the properties of an object match up with those properties adverted to in the appropriate concept. In the second way, called the 'non-ascriptive mode', our reference is fixed in virtue of their being a relation of causal co-variation between the properties of objects in the world and the character of our experiences. (This is exemplified by the forms of representation employed in the writings of Dretske and Tye for instance, as mentioned above<sup>14</sup>). The important feature of non-ascriptive modes is that we need not be aware of the relation between the representation and the object represented; it is sufficient that it be the case that the two are causally linked.

Now it is possible to deal with problems caused by the conceivability argument. There is an object that we refer to ascriptively, according to some physical property, such as being an instance of C-fibre activity. We can also refer to it non-ascriptively, in virtue of some qualitative property, such as painfulness. The original demand of the conceivability argument, was that these two properties had to be contingently related: otherwise, it would be a matter of necessity that they were linked, and yet this necessity does not show up when we think about the two properties. However, this demand can now be side-stepped, by differentiating between the two modes of presentation. C-fibre activity and painfulness are related by necessity; the fact that the relation appears to be

---

<sup>14</sup> For a discussion of Representationalism, see Chapters 3 and 4 above.

contingent, is due to the feature of non-ascriptive modes of presentation - that the relation between the representation and the object represented, is one that holds without our being aware of its existence.

### Subsection 2 : Levine and the 'explanatory gap'.

The above debate has resulted in a Pyrrhic victory for Materialism. The introduction of non-ascriptive modes of presentation, may offer some respite for the Materialist, but it remains to be proved that qualia are necessitated in the way suggested above. As the matter stands, there is a gap between how the world reaches up through our senses into the brain, and from there into qualia. The defence of the previous subsection, shows at most that Materialism is not ruled out as an option; but some form of Substance Dualism is just as consistent with these conclusions. It is interesting to see that Levine mentions a stand-off between the Materialists and Dualists that arises from the above:

The main burden of the physicalist argument is borne by considerations of causal interaction. If qualia aren't physical processes (or realized by physical processes), then it becomes very difficult to understand how they can play a causal role in both the production of behaviour and the fixation of perceptual belief.<sup>15</sup>

This is the very same dilemma that arises in McGinn's discussion of the nature of the mind; we might not ever know that it is physical, but our faith in mental causation can only be maintained if we assume that the mind is part of a single physical order. Yet the mystery of interaction that is supposed to vex the Dualists, is mirrored by a mystery of embodiment that is currently embarrassing the Materialists.

The problem lies with there being no satisfactory explanation why physical creatures should experience the world in the manner that they do. Given any putative identity between the reference of a physical description and a particular mental state, it is conceivable both that the description could apply in the absence of that mental state, as well as the mental state could obtain even though the physical description did not hold. This amount of flexibility in explanation would not be tolerated in any theory, and the Materialists should not be permitted to be an exception. So why is there a problem with providing a Materialist account of mental types - qualia, for instance?

Levine suggests the following. Though the nature of explanation is far from clearly understood, it is reasonable to assume that we expect an explanation to uncover causal relations between properties. Initially, we notice superficial patterns of behaviour through perception, couched in a primitive scientific vocabulary; for instance, that plants tend to grow towards the sun. This leads us to attribute causal properties to the sun with respect

---

<sup>15</sup> Levine. 1993. P 126.

to the plant. In order to advance our understanding of these properties, we investigate the processes that lie behind this phenomenon. What we find includes the presence of growth hormones in the plant's tissue, that are inhibited by light. This results in an imbalance of hormone on the shaded side of stems. Growth is therefore slower on the illuminated side, resulting in the stem bending towards the light source. In articulating this account, many other theories are introduced that are not solely connected with plant growth; such as concern the nature of energy, light waves and Natural selection for example. This account of explanation being a two stage process, is akin to Nagel's view of how we come to formulate theories.

If the above analysis is accurate, then there is a problem for explaining such mental properties as qualia. The problem is in a sense twofold. Not only is it impossible to think of qualia except as they appear to us, but we have no clear reason for identifying any particular appearance type with any particular causal rôle. That this is evident, is confirmed by 'conceivability arguments' as well as the possibility of inverted spectra. The particular quale I associate with red objects may cause me to have certain beliefs about the world, but it is not clear that a different quale associated with the same natural properties could not do just as well. Whilst the nature of qualia seems to be exhausted by their appearance, there does not seem to be any definitive limit to the causal rôle they might play; as such, their causal properties are seemingly boundless. If we proceed in our understanding of the world by passing from appearance to reality, by uncovering the *causal* properties of phenomena, then qualia are ill-suited to being the subjects of our powers of comprehension.

When we attempt to explain what it is about a physical substance or process that causes a particular quale, we find that our understanding falters: there is a gap between the physical and the mental which methods of explanation seem unavoidably incapable of bridging. Levine sums up as follows:

What seems to be responsible for the explanatory gap, then, is the fact that our concepts of qualitative character do not represent, at least in terms of their psychological contents, causal roles.<sup>16</sup>

Materialism cannot simply shrug this problem off. It is a poor feature of any theory that it cannot explain the nature of a vast and significant area of our mind. Of equal concern is the implications this result has for our cognitive powers. Granted that 'understanding' is a psychological process, there has to be a reason why the gap exists at all; and this reason has to advert to some physical feature of our cognitive system. The Materialist owes us an explanation of why a physical process inevitably has these lacunae; the next section outlines McGinn's attempt to provide such an account.

---

<sup>16</sup> Levine. 1993. P 134.

### Section Three

McGinn develops his defence of Materialism against the above claims, in a series of books and articles: the following is a brief summary of the arguments as they bear upon the present issue. There are two important aspects to his position: one relates to his view of philosophical problems generally; the second to the problem of the relation between the mind and the body in particular. In "Problems in Philosophy", he argues that such problems are generated by limitations in our cognitive abilities, and not because of the objects with which philosophy is concerned. There is nothing mysterious about the nature of reality:

Reality itself is everywhere flatly natural, but because of our cognitive limits we are unable to make good on this general ontological principle. Our epistemic architecture obstructs knowledge of the real nature of the objective world. I shall call this thesis *transcendental naturalism*, TN for short.<sup>17</sup>

The thesis concerns different types of question, and different types of mind. A particular question 'Q', may be thought of as 'problematic' or 'mysterious', depending upon the type of mind that is attempting to provide an answer. 'Problems' are those for which the mind has, in principle, the cognitive ability to solve; whereas 'Mysteries' are those for which the mind has, in principle, no such cognitive ability. It is possible therefore, that there should be some questions which are mysteries to one mind type, but only problems to a different mind type.

McGinn argues that there is no metaphysical difficulty involved in discovering what links the body with consciousness. For our types of mind, the question is a 'mystery'; it is conceivable that there are mind types for which the question is simply a 'problem'. Where there is no metaphysical difficulty, there is no philosophical interest. In that sense, there is no philosophically interesting mind-body problem; there is only a deficiency in our cognitive capacity. Once philosophy has established this conclusion, there is no more work for it to do. Regarding the brain as the seat of consciousness is therefore a viable option. So far as McGinn is concerned:

The problem arises...because we are cut off by our very cognitive constitution from achieving a conception of that natural property of the brain (or of consciousness) that accounts for the psychophysical link. This is a kind of causal nexus that we are precluded from ever understanding, given the way we have to form our concepts and develop our theories.<sup>18</sup>

The arguments that support this thesis may usefully be considered as falling into three parts. The first part defends the claim that there is a property of the brain that is responsible for the linking of consciousness with physical matter. An outline of this defence is provided in Subsection 1. The second part argues that the

---

<sup>17</sup> McGinn. 1993b. Pp 2-3.

<sup>18</sup> McGinn. 1991. P 2.

nature of this property is beyond our cognitive capacities. This argument will be presented in stages, in Subsections 2 to 5. Finally the third part concludes that the problem is not one for philosophy to solve, it is just that we have to learn to live within our cognitive limitations. The following is a short presentation of the above, before going on to consider problems with this thesis in Section Four.

### Subsection 1 : Why the psycho-physical link is a natural property of the brain.

In order to explain the psycho-physical link, McGinn assumes that some form of Materialism is true. One might think that it is odd that McGinn takes himself to be a Naturalist, however Transcendental. His position is made consistent by the distinction between two types of Naturalism - 'effective' and 'existential'. The former makes the following strong claim concerning the link between ontology and epistemology - what exists, is in principle knowable by us. (As Levine notes<sup>19</sup>, this is a claim for which there is as yet no argument) He rejects this version as being wholly unmotivated, and far too sanguine about our abilities. Instead, he describes himself as holding a weaker thesis - existential naturalism - as follows:

...nothing that happens in nature is inherently anomalous, God-driven, an abrogation of basic laws - whether or not *we* can come to comprehend the process at work.<sup>20</sup>

He also assumes that Eliminativism is mistaken. These assumptions are justified partly because the alternatives are too horrible, or too absurd, to contemplate; partly because we are familiar with many correlations between brain states and mental states, suggesting that such a link exists; and partly because consciousness may be regarded as being an extension of the development of life. Only the third consideration is in any sense an argument. The idea is that if we are happy to believe that 'life' evolved from inorganic matter, then we should have no qualms about 'consciousness' doing so; for consciousness may be thought of as a natural consequence of the evolution of living things. Just as we are developing a theory to explain how aggregations of inert material become alive, so too, in time, may we come to understand how organisms developed a property that mediated between physiological structures and consciousness. The mistake is to assume that we can always provide such an explanation.

### Subsection 2 : Cognitive closure and the CALM hypothesis.

The second section of the thesis, is the most complex. However, it is important to note that it assumes the truth of the first

---

<sup>19</sup> Levine. 1993. P 124.

<sup>20</sup> McGinn. 1991. P 87.



section; that is, it is assumed that there is a physical property of the brain that is responsible for consciousness. Instead of providing an account of this property, McGinn intends to demonstrate that we can never know its nature. In order to do this, he introduces the notion of 'cognitive closure'<sup>21</sup> and the 'CALM'<sup>22</sup> hypothesis. These refer to features of human cognition; in particular, the ability to manipulate concepts, and the ability to form theories. Our abilities in this respect are taken to be a result of Natural Selection. In some sense, it is quite accidental that we seek knowledge the way we do. However, the fact that humans have had to struggle to survive, supports the analysis that McGinn offers. It is interesting to note how this chimes in with Jackson's comments at the end of his article on qualia, although the two philosophers have opposite views upon the nature of consciousness.

Cognitive closure is the relation a type of mind has with respect to properties, and theories concerning those properties. By definition, a mind type 'M' is cognitively closed with respect to property 'P' or theory 'T', if and only if, the concept-forming abilities of 'M' are insufficient to comprehend 'P' or understand 'T'. The moral of this idea is that there is no warrant for concluding something does not exist simply on the basis that we cannot form a conception of it. This suggestion is supported by instances of perceptual closure, where animals are not sensitive to the same range of perceptual experiences as others; so for example, frogs are perceptually closed to the colour properties of objects, since their visual system has no capacity for discriminating colours. However, we would not conclude the non-existence of colour properties on the basis that frogs cannot detect them. Similarly, some minds may be unable to form theories that account for some events; it would be reasonable to suppose that ants are incapable of understanding the structure of sub-atomic particles, for instance. In this case we should say that such theories were cognitively closed to ants. What is important, is that one may be cognitively closed with respect to a theory, but still be in a position to appreciate that a gap exists in one's understanding. Our position with respect to the mind-body problem, is that we know that there must be a property that links the two, it is just that we are cognitively closed to that property.

The 'CALM' hypothesis supplements the above idea. It is a positive account of what human brains are good at doing in the way of problem solving. 'CALM' stands for: "combinatorial atomism with lawlike mappings". When faced with questions, human brains typically produce answers by building up from a store of basic elements that combine together in certain pattern-like ways. McGinn cites the disciplines of Physics, Linguistics and Mathematics, as providing paradigms of such methods of dealing with problems. Understanding any particular feature of the world, is achieved through placing it within this network of elements and laws; such as when the properties of gold are understood by reference to its composition and how it is related to its environment, according to

---

<sup>21</sup> McGinn. 1991. Pp 3-5.

<sup>22</sup> McGinn. 1993b. Pp 18-20.

observed regularities; so it is not surprising that it does not rust, fails to keep a sharp edge, and feels heavy when handled.

We are faced with the following question: which property of the brain accounts for consciousness? McGinn's response is that there is such a property 'P' - it is just that we are cognitively closed to 'P'. Given that 'P' connects the brain with consciousness, it might be thought that its nature would be revealed either by studying consciousness via introspection, or the brain via perception. However, on reflection, neither method proves satisfactory.

Subsection 3 : Attempts to discover the nature of the link via introspection.

Consider methods of introspection. So far as revealing properties of the brain is concerned, introspection is clearly not up to the task. No matter how hard I introspect my conscious states, I will not thereby be able to discover properties of the brain. In fact, this method is hopelessly unproductive of knowledge, when compared with perception. One might think that it was a faculty that was sensitive to conscious states; however, it would be more accurate to say that it is only sensitive to a small sub-class of those states - those conscious states that are occurring to me. The conscious states of others are not discoverable by my powers of introspection. As a consequence of this, it is unable to provide instances of experiences to which I am not normally sensitive. For example, it is impossible for those born with defects of the senses - such as hearing - to form accurate conceptions of what the sound of bird song is like, on the basis of introspection. These shortcomings in introspection, make it plausible that not all properties of consciousness are known to us.

The idea that there is a hidden structure to consciousness is important in the overall argument, for it allows that the property 'P' is equally a property of consciousness even though we are not aware of it. Therefore, there have to be some properties of consciousness that are hidden to us. If we assume that all of consciousness is open to introspection, then we are likely to deny that such a property 'P' exists, since introspection cannot discover it. However, the above considerations suggest evidence of a confusion between infallibility and completeness. McGinn is not querying the infallibility of introspection, but rather he is saying that its scope does not exhaust the properties of consciousness. Independently of the rôle it would play in supporting the 'hidden' link between mind and brain, there is some reason for accepting the idea of a concealed region of consciousness; without this notion, the following would be inexplicable: the logical properties of conscious thoughts; and the condition known as 'blindsight'.

Taking the above one by one, there is a relation between the apparent meaning of a sentence and its underlying logical structure. In order for sentences to stand in logical relations to one another,

there must be some forms upon which these relations depend. The nature of these forms is not revealed at the level of appearance, for if we are guided by the superficial appearance of sentences, then we are too easily led to paradoxes and inconsistencies. These are resolved by appeal to the theoretic logical form of the sentence. Insofar as our conscious thoughts are taken to be analogous to sentences, then they too must have an underlying form which is distinct from their appearance to introspection. When we reflect upon the nature of thought aided solely by introspection, it is no surprise that there are problems analogous to those apparent in sentences prior to logical analysis: for instance, such things as holding two or more beliefs that are inconsistent with one another, or failing to acknowledge a belief even when it is entailed by others that we hold. To dissolve the problems inherent in thought, we have to get beyond their surface properties; therefore, just as there is a deeper level of language, there must be properties of consciousness that are hidden to introspection.

If one could show that perceptual processes are possible, even when properties of consciousness open to introspection are absent, then one has good reason to suppose that there are further mental properties upon which the processing depends. 'Blindsight'<sup>23</sup> is a condition that suggests that some conscious processes are outside the range of introspection. It is a condition whereby individuals lose sensitivity to part of the normal field of vision. However, when articles are placed within the affected range, although the individuals claim that they can see nothing, when asked to guess whether or not the object is present, they are correct more often than is consistent with typical guesswork. One explanation for their accuracy, is that some processing is taking place, even though it does not include the generation of visual experiences. McGinn takes this to be further evidence of conscious processes taking place that enable us to act upon information about the world, without those processes being revealed to introspection.

#### Subsection 4 : A 'reductio' to prove that the link is essentially hidden.

To re-cap, the above considerations were intended to show that the faculty of introspection has limited utility with respect to discovering facts about ourselves and the world. In addition, there a number of properties of consciousness that are hidden to this faculty; it is assumed that 'P' falls within this set. There is no reason as yet to assume that 'P' is an essentially hidden property. However, McGinn has an argument which draws this conclusion.

The argument takes the form of a 'Reductio ad Absurdum'. It assumes that we possess knowledge of how it is that specific forms of consciousness arise from processes and structures in the brain. As a result of this, our theories of consciousness should generalise to all types of consciousness, including those of bats for instance. In particular, consider a bat experience type 'E<sub>b</sub>', which our theory

---

<sup>23</sup> For more on Blindsight in relation to Epiphenomenalism, see Chapter 7, Pp 166-167.

tells us depends upon neural property 'P<sub>b</sub>'. By understanding the nature of 'P<sub>b</sub>', we should be in a position to appreciate how the bat's brain generates 'E<sub>b</sub>'. That is to say, we would understand both the physical nature of 'P<sub>b</sub>', as well as the experiential nature of 'E<sub>b</sub>': so we would know what it was like to be a bat. This line of reasoning results in a dilemma. Either we can grasp this theory, in which case we will be aware of the qualitative aspect of 'E<sub>b</sub>'; or we can not grasp this theory, because we cannot be aware of 'E<sub>b</sub>'.

It is already assumed that the formation of experiential concepts depends upon the range of experiences available to the individual. Accordingly, where an organism lacks the structure that underlies a particular experience, it will be unable to form a concept of the character of that experience. In the case of humans, it is therefore impossible to form concepts of bat experience. Hence, the experience 'E<sub>b</sub>' is closed to us; and in virtue of the dilemma above, so too is the theory that accounts for the relation between mind and brain. So, if we assume that we can form a theory to account for the way consciousness is produced in the brain, then we end up with the absurd conclusion that it is possible to share the experiences with other creatures; this conclusion is absurd, since it contradicts the principle that we cannot form concepts of experience outwith those connected to our senses.

McGinn concludes that:

...our concepts of consciousness just *are* inherently constrained by our own form of consciousness, so that any theory the understanding of which required us to transcend these constraints would *ipso facto* be inaccessible to us.<sup>24</sup>

#### Subsection 5 : Attempts to discover the nature of the link via perception.

If introspection is an inadequate method of understanding property 'P', then perhaps the faculty of perception will be successful? McGinn argues that perception is equally unsuitable. There is nothing about the sight of the brain that immediately suggests that it is the organ responsible for producing conscious states. It might be added that the more we learn about the brain, the more improbable it seems that any structure could serve as 'P'; for after all, as mystery about the brain diminishes, there are accordingly fewer places where 'P' could be lurking. However, it is not that we might come across 'P' by some unanticipated stroke of good fortune; we are inherently disqualified from making such a discovery.

This bold claim depends upon the relation between the faculty of perception and the range of concepts that can result from its use in the formulation of theories. McGinn alludes to two features of concept introduction to support his contention that we will wait in vain for the arrival of a comprehension of 'P'. The first is the observation that our senses pick out spatial properties of objects.

---

<sup>24</sup> McGinn. 1991. P 9.

Hence, when directed upon the brain, they are sensitive solely to the arrangement in space of various sorts of tissue and processes. However, the nature of consciousness is not completely intelligible in spatial terms; for instance, how would one describe the spatial aspect of a piece of reasoning, or the feeling of elation? Whenever a process were championed as being responsible for creating these cognitive effects, it seems that it would always be open to doubt that the claim was true.

The second feature of concept introduction utilised by McGinn, is the 'principle of homogeneity'<sup>25</sup>. This principle governs which concepts it is legitimate to employ in the formation of theories; basically, it proscribes the introduction of mental concepts into physical theories and vice versa. It effectively rules out the possibility that 'P' will appear in meaningful hypotheses, in virtue of its being, in part, a mental concept. McGinn anticipates the following objection; just because we cannot sense 'P', it does not mean that we are unable to form an understanding of such a property. For indeed, if we were limited to what we could sense in the formulation of theories, then we would have a very different comprehension of the world from what we possess at present. Thus, the fact that we cannot observe the link is no reason for thinking that we cannot bring it within the scope of our knowledge; after all, we have a grasp of the concept 'electron', for example. Why cannot some form of inference to the best explanation be used here?

The difference between 'P' and 'electron', is that our understanding of the physical world was incomplete before the introduction of the term 'electron'. There were all sorts of occurrences which could not be explained upon the basis of existing theories. These events, such as traces in a cloud chamber, could be best explained by the action of physical particles with certain properties. With the addition of this concept - electrons - to our theories, our ability to explain events in the world became more comprehensive. In the case of the brain, this strategy may be successfully employed, but McGinn suggests that our understanding of the brain is not improved by introducing concepts of consciousness. This is a view that may also be found in Nagel<sup>26</sup>, that the physical matter of the brain has a *purely* physical explanation. Since we do not have any concepts of the linking properties of consciousness, it is clear that the introduction of mental concepts will be useless in our construction of physical theories.

Besides, theoretic concepts are generally formed by extending the concepts we already have, based upon observation. Thus, for instance, the concept of a 'molecule' may be thought of as arising from the application of the relation of composition - such as holds between a house and its component parts - to smaller and smaller degrees. This strategy in the end can only produce concepts whose observation-based parentage cannot be eliminated; since consciousness in the brain cannot be discerned by observation, then

---

<sup>25</sup> Attributed by McGinn to Nagel in Panpsychism. Nagel 1979. P 183.

<sup>26</sup> Nagel. 1979. Panpsychism.

none of the concepts derived from observation will be adequate for 'P'.

#### Subsection 6 : Learning to live with 'Cognitive Closure'.

McGinn ends by pointing out that it should not be particularly surprising that we lack the cognitive resources to comprehend how it is that matter can cause conscious experience. Neither introspection nor perception individually, or in conjunction with our ability to form theories, are suited to discovering this link. Ironically, ignorance of this property is a consequence of our particular cognitive profile. To be acquainted with the nature of 'P', would probably require the ignorance of some other types of property. So on balance, we may consider ourselves relatively fortunate. Knowledge of 'P' will not help us to survive, so why expect to have an evolved brain capable of providing that information? Our concept forming abilities are well suited to enabling us to survive; their singular limitation is frustrating solely to philosophers.

#### Subsection 6 : Conclusion.

The first two parts of his argument substantiate the claim that there is no philosophic problem about the link between the mind and the body. There has to be a natural link, for reasons outlined in the first part. This effectively takes care of the ontological aspect of the mind-body problem. From the second part, it is clear that there is a natural property of the brain, to which our minds are cognitively closed. The mind-body problem has been generated by the thought that, in principle, it is always possible to have full knowledge of everything that exists. McGinn has provided substantial grounds for doubt in this proposition. Once we break away from the restraint of this belief, then we can flee the philosophical tyranny of this puzzle.

### Section Four

The problems with this position fall into four broad categories. They do not all have the same force; some are clashes of intuition which might make TN less plausible; some are, I think, mistaken; but ultimately, they strongly suggest that a Materialist cannot take refuge in this sort of strategy as McGinn presents it.

#### Subsection 1 : Is McGinn's Naturalism compatible with his Agnosticism?

The first category of criticism focuses upon the problem of being a Naturalist and being an agnostic. Both Fodor and Dennett accuse McGinn of underselling the virtues of Materialism; after all, the idea that consciousness is a property of the brain, has gained credence because of the success that science has enjoyed in explaining our world. Our faith in Materialism is partly based on the

promise that science will banish ignorance. Yet here we must ask ourselves: if science is going to renege on that promise when applied to 'consciousness', then why should we believe that the mind-body problem is only a scientific as opposed to a philosophic matter? Fodor suggests that McGinn should evince a greater sense of failure, for it is a vacuous form of Materialism that stops short of dissolving the mind-body problem:

What we imagine to be compatible with our materialism is, I think, just about the same as what we can imagine ourselves being able to explain. That consciousness might be one but not the other strikes me as a wan doctrine and an empty hope.<sup>27</sup>

Dennett<sup>28</sup> similarly accuses McGinn of inconsistency. Having embraced a Materialist position, there is no point in abandoning the expectation that answers will be forthcoming. Certainly there is a connection between the way we are constituted which makes understanding the nature of the link difficult, but this is because we are attempting to 'reverse engineer' the brain in an attempt to comprehend how it works. This strategy is used to uncover the nature of the properties of an object, by employing our understanding of engineering, to work from the finished product 'O' back to how its constituent parts would need to articulate in order to explain why 'O' possess its various properties. However, our present engineering techniques are based upon isolated units of functionality, which are simply not applicable to the workings of the brain as we understand them to date. Yet there is no reason to suppose that we will never come to adapt our engineering concepts in such a way as to be able to comprehend the complexity of the brain.

It is difficult to know how seriously to take these objections. There is a sense of disappointment attendant upon arriving at the New Mysterian's conclusion, but that may be because we had built up our hopes unrealistically. Against Dennett's point, McGinn might agree with the diagnosis of the difficulty - it might even lend some support to his argument - but there is perhaps still room for him to deny that Dennett has any grounds for optimism. In short, these considerations fail to provide sufficient grounds for abandoning McGinn's position.

### Subsection 2 : Does McGinn make unreasonable demands of science?

The second category of criticisms focuses upon McGinn's argument against our ever knowing the nature of 'P' through introspection. Both Kirk and Flanagan suggest that the notion of 'fully explain' or 'fully grasp' is being used inconsistently with what we normally accept as a theory providing a 'full explanation'. Both note that as used by McGinn, a theory fully explains the nature

---

<sup>27</sup> Fodor. 1991b. P 12.

<sup>28</sup> Dennett. 1991. P 273.

of 'P' if it would allow us to know not only that 'P' gives rise to certain experiences, but that such knowledge would also enable us to have those experiences. This debate is of the same kind as raised by Jackson with the Mary example: if we knew everything about the perceiving of colour, then we should know what it is like to perceive colour. Flanagan<sup>29</sup> suggests that this is asking too much of a theory; no-one supposes that theoretical knowledge will suffice for actual experience. Nor does he see that we gain much from being able to share experiences. For instance, we now realise that bats are not blind: knowledge of the types of experience available to bats has increased regardless of our not being acquainted with the character of these experiences. These issues are also raised in conjunction with problems concerning the 'Homogeneity constraint'<sup>30</sup>, which similarly places the standard of scientific rigour at unrealistically high levels.

How damning these arguments are, depends in part upon how much one expects of a successful theory of consciousness. Against Flanagan, a possible response would be to insist that we are still left with a brute correlation between brain state and experience. There is still room for alternative views upon the nature of mental substance. No-one is contesting the correlation between brain states and psychological states; what is undecided is whether or not this is a correlation between two different types of substance. So long as there is no Materialist theory of how matter gives rise to experience, the findings of science are consistent with Substance Dualism; and this is the situation that McGinn and Jackson attend to, and their critics blithely ignore.

### Subsection 3 : Problems with McGinn's 'reductio' argument.

In "Why Shouldn't We Be Able To Solve The Mind-Body Problem?", Kirk offers an argument against the assumption, 'A', that McGinn makes concerning the existence of the property 'P' and theory 'T':

(A) There exists some property *P*, instantiated by the brain, in virtue of which the brain is the basis of consciousness. Equivalently, there exists some theory *T*, referring to *P*, which fully explains the dependence of conscious states on brain states.<sup>31</sup>

The problem lies with how to interpret the expression "fully explains". For McGinn, it means that anyone in possession of 'T' will thereby have access to the character of all possible experiences. Kirk prefers a different interpretation of "fully explains". This allows that possession of the theory will enable us to appreciate the functionality of the experiences, and how they are subserved by the brain - but not what those experiences feel like. McGinn is committed to the existence of 'T'; Kirk argues that the nature of 'T'

---

<sup>29</sup> Flanagan. 1992. Pp 94-97.

<sup>30</sup> Flanagan. 1992. Chapter 6. Pp 109-128.

<sup>31</sup> McGinn. 1991. P 6.



is unintelligible. Essentially: "No possible theory, accessible or not, could do what he assumes T must do."<sup>32</sup>

Kirk's argument runs as follows. Any naturalistic theory will proceed stepwise from what we know about the physical properties, to the point at which we can see how such properties make sense of particular experiences. This process depends upon there being a prior acquaintance with the experience that the theory is explaining - where this acquaintance is made directly or via one's imagination. However, consider the case of the bat's experiences. McGinn's theory 'T' should explain why the physical property 'P<sub>b</sub>' underpins the particular bat experience 'E<sub>b</sub>'. However, in order to arrive at this theory in the first place, it seems we must have some prior knowledge of 'E<sub>b</sub>' before we are able to take the final step: and this we do not have, either through experience or act of imagination. So in order to arrive at 'T', we would have already to possess knowledge of what 'E<sub>b</sub>' was like; and this was supposed to be a result of 'T' rather than a pre-requisite to forming 'T'. The problem for McGinn, is that he appears to be committed to there being a theory 'T' that will enable us to appreciate the character of 'E<sub>b</sub>'. But no theory can do this without prior acquaintance with that experience. So 'T', as characterised, is impossible to formulate. Therefore the assumption 'A', that is crucial to the rest of McGinn's agnosticism, is false insofar as it involves an impossibility.

There is a possible defence against this objection. Much depends upon the powers a theory is expected to confer upon anyone in possession of it. Considered as types, theories are the wrong sorts of things to provide sensations. Certainly, tokens of a theory may be experienced as sounds or sights - whether real or imagined - but this is not what is meant by the idea that 'theories enable one to grasp the character of experiences'. What the theory envisaged must allow, is an understanding of what the properties of physical matter are, that produce sensations when stimulated in certain ways. Any Materialist must assume that, in principle, there is a potential explanation for why sensations have the character that they have; and that this explanation is in terms of the chemical properties of structures in the central nervous system. There must be some quality that brain tissue has, that determines with which of the myriad possible sensations it is associated. If we knew how sensation was related to different structures of matter, there seems every reason to suppose that that we should be able to tell which type of experience went with which type of physical organisation. Analogously, knowing how the structure of a molecule determines its behaviour, enables us to predict how certain molecules would behave, even though they are absent from our environment.

It is accepted that theory construction works in a step-like manner. The criticism is, that so far as the bat experience 'E<sub>b</sub>' is concerned, we require a prior knowledge of it before taking the final step that relates the physical with the mental. However, this is to misrepresent the nature of 'T'. 'T' is supposed to be a theory that relates physical properties with conscious ones. Insofar as I am

---

<sup>32</sup> Kirk. 1991. P 17. From McGinn 1991 P 6.

acquainted with my own experiences, there does not seem to be any obstacle to forming 'T' with respect to my own case. Thus there is no problem making the step-like investigation from brain states to first person reports of my consciousness. It is reasonable to assume that it should be possible to generalise theories from the particular to the general. Having devised a theory that relates physical to mental states in my case, it should be possible to extend this theory to all types of consciousness, since they are all subserved by the same types of material. That being so, we should expect to be able to know the quality of a bat's experiences as easily as we know those of another human. The impossibility of formulating 'T' only arises if it is assumed that the theory of human 'P' does not generalise to other mammals, and physical objects generally; and this assumption is damaging to the applicability of scientific practice.

Having said that, there is a problem with this argument. Crucial to its validity, is the acceptance of the idea that our concepts are constrained by our experience. The *Reductio* works only if the conclusion is false; the conclusion being that theory 'T' would enable us to grasp such things as bat-type experiences. However, the only reason we have for regarding this proposal as false, is based upon our *current* beliefs concerning the extent to which it is possible to share the feelings of others. When we assume the existence of such a theory, 'T', it must alter our preconceptions on the matter. If I were blind from birth, and yet knew how it was that matter subserved experience in the way that McGinn proposes his theory would, then I would thereby know what it was like to see things; there would be no such constraint upon my concept formation. We only have reason to limit our concepts in the way he suggests because we lack the theory 'T'. Once in possession of 'T', there seems to be no reason for enforcing such a constraint. In which case, the absurdity does not arise. I do not think, therefore, that McGinn has successfully shown that we are cognitively closed to such a theory.

#### Subsection 4 : Does McGinn's position redeem Materialism?

The most telling objection to New Mysterianism, is that it is principally an argument for 'cognitive closure', rather than Materialism. The defence of Materialism is largely intuitive. An argument was singled out in the beginning, to the effect that consciousness was like life in terms of its development, and so there was every reason to believe that, like life, there was an explanation of consciousness in natural terms. However, McGinn notes a significant dissimilarity between the two in the article 'Consciousness and the Natural Order.' He writes:

Consciousness could vanish from the Universe for ten minutes and then reassert itself after the interlude, so long as the right reproductive processes

had been initiated. But if matter or life vanished for that period, there would be a devil of a job getting them started again.<sup>33</sup>

In this sense, the generation of life and consciousness are quite distinct. Whether or not this means that consciousness is not material, it certainly weakens any support Materialism is supposed to have, as being the obvious ontological choice.

In many ways, the above quoted observation could equally be made by a Substance Dualist. Here is where the weakness of TN lies. If you are pre-disposed to Materialism, then it provides an alternative stance to overcome the sorts of difficulty raised earlier - by Jackson, by Kripke and by Levine. If you are open minded, then the philosophic problem re-emerges with the initial arguments for TN. The notion that the cognitive closure arguments could be run for a defence of Dualism, is not lost on McGinn. In 'Consciousness and Cosmology - Hyperdualism Ventilated', he imagines a dialogue between himself and an extra-terrestrial Dualist, who believes that there are two universes - U1 and U2 - which interact with each other; that mind is a product of the abstract, and that there is 'ontic harmony, if not penetration' between the two universes. He notes that what whereas 'embodiment' is the problem for Materialism, it is 'interaction' so far as the Dualist is concerned: and just as Materialism can avoid awkward questions about embodiment by giving the cognitive closure arguments, this is precisely the same strategy that is open to the Dualist concerning problems with interaction. Indeed, he has his interlocutor admit:

Ironically enough...our brains may be preventing us from achieving a scientifically powerful theory of the deep workings of U2, and indeed of the manner of its connection with U1...[the nature of U2] might be quite unknowable to the kinds of being who benefit...from its manifestations.<sup>34</sup>

The arguments about cognitive closure, even if valid, do not provide a distinctly Materialist position. Materialists may deny that they have an obligation to explain how the properties of mind are fundamentally physical, because the link between mind and brain is essentially unknowable. However, for this move to be plausible, it must be for reasons that make Materialism a viable option in conclusion, rather than in assumption. If it is possible to discharge this obligation in the above manner, then it will have to be approached from a different direction. It is the aim of the next two chapters to indicate that approach.

---

<sup>33</sup> McGinn. 1991. Pp 46-47.

<sup>34</sup> McGinn. 1993a P 173.

## Chapter Six

One may be forgiven for thinking that this thesis is Anti-Materialist in outlook: the previous five chapters have been written with the sole intention of pointing out the inadequacies of the leading Materialist positions. If they have been successful, they place the author in the embarrassing situation of having to support something that he has hitherto been at pains to show in its weakest light. Regrettably, the time has come to leave the comparatively painless task of criticising Materialism; this chapter must argue for the possibility of a Materialist solution to the 'Mind-Body' problem.

The position that is being argued for, may be summed-up as follows. States of consciousness are caused by properties of brain tissue; the nature of these properties is essentially beyond our understanding. Purely for the benefit of clarity, this position will be referred to as 'Agnostic Materialism', or 'Agnosticism' for short.

I see no point in doubting the existence of relations between physical and mental states. The question to be addressed, concerns the nature of these relations. There are two facets of the relation between the mind and the body that have posed the following questions: does the relation lie within an ontological unity, or between distinct ontological types: is the relation one of sheer coincidence, or is there a law-like causal link between the two? These questions are independent of one another, insofar as either answer to the first question is compatible with either answer to the second. Therefore they require separate consideration. This chapter and the next will concentrate upon defending the following positions; that there is a relation within a single ontological type, and that the relation is a law-like causal connection between the mind and the body.

The argument in support of the first answer owes much to the comments made by McGinn. As such it seeks to justify an agnostic stance on the basis of limitations to our understanding. If these arguments are cogent, then our ignorance of how the mind could be realized by physical properties will not tell against Materialism. It will be the purpose of this chapter to present those arguments. Briefly, the strategy is to establish the idea that the incompatibility between mental states and scientific reduction does not have to be explained in terms of a difference in substance. Rather, it can be accounted for in terms of a relation between the way we understand the nature of things, and the set of objects available to such processes. Although substantially the same as other phenomena, considered as objects of scrutiny, these processes are themselves excluded from that set. This results in the two distinct ways of comprehending ourselves and the world; respectively the holism of Psychology and the empiricism of Physics.

Section One will provide some background to the notions of 'cause' and 'explanation' that are to be employed in what follows. Section Two re-examines the remarks made by Kripke, alluded to in

the previous chapter<sup>1</sup>, in an attempt to defuse their anti-Materialist consequences. Sections Three and Four suggest that our ignorance of the ontological nature of the relation is a corollary of our processes of information gathering. Once this defence is in place, granted the debt it owes to McGinn's position, it is important to analyse the reasons for the failure of New Mysterianism, in order to avoid leaving this account open to a similar objections. Thus Section Five will briefly argue that this form of Agnosticism lends itself exclusively to Materialism.

The argument in support of the existence of law-like causal connections between the mind and the body, depends in part upon the success of this chapter. That such connections exist is, in effect, a denial of Epiphenomenalism. The reasons for this denial will be given in Chapter Seven.

### Section One

When the main focus of one's attention is the relation between the Mind and the Body, it is inevitable that other issues in philosophy will be presumed to have greater consensus than is warranted. The constraints of space determine that only a barely adequate account be given of topics more intelligently appreciated by Philosophers of Science. In good faith, it is assumed that where there is still division over how certain concepts are to be understood, the use to which they are presently employed is not wholly outrageous.

#### Subsection 1 : A brief introduction to the terms 'cause' and 'explanation'.

In his book "The Facts of Causation", Hugh Mellor identifies two basic sorts of causes and effects. On the one hand, there are true states of affairs; for instance, "The plant withered", and "The plant was exposed to frost". These can be related as cause and effect as follows; "The plant withered because it was exposed to the frost". On the other hand, there are relations between particulars (which correspond to names and other referring expressions), which might be either things or events. Particulars, such as the withered plant and the frost, can be related thus; "The frost caused the plant to wither". He goes on to argue that 'facts' in the above sense, constitute the fundamental sort of causal relata.

In everyday discourse, it is taken to be a general feature of causes, that they must somehow necessitate or determine their effects. We suppose a cause 'a' to be necessary for the effect 'b', when we believe that if 'a' had not occurred then neither would 'b'. Similarly, we take cause 'a' to be sufficient for its effect 'b', when we believe that if 'a' had happened, then so would 'b'. However, Mackie<sup>2</sup> demonstrates that this interpretation of causation is too superficial. There are cases where a state of affairs is caused by a

---

<sup>1</sup> Chapter Five Pp 116-117.

<sup>2</sup> Mackie. 1986. Chapter 3.

concatenation of circumstances, none of which are individually necessary nor sufficient for the effect, but which nonetheless significantly contribute to the effect. Mellor construes the causal relation in terms of a cause making the effect more or less likely. Finally, there is an assumption that there are laws of nature, which he considers to be broader than simply scientific laws, to relate causes to effects in a manner which is more than simply accidental.

So far as the nature of 'explanations' is concerned, contemporary discussion begins with the work of Hempel<sup>3</sup>, and his presentation of two influential models of explanation. The 'Deductive Nomological' model (or D-N, for short), treats explanations as forms of deductive inference, from a major and minor premise (or premises) to a conclusion. On this model, the phenomenon to be explained (the 'explanandum') occupies the place of the conclusion, and the premises are taken by the propositions that are adduced to provide the explanation (the 'explanans'). Typically, the explanans is formed by the conjunction of a proposition describing a single instance, with a proposition derived from a law of nature. For instance, the explanation for why lamps never work so well after the inside of the light socket has been dipped in plastic, would run as follows:

1. The terminals of this lamp are covered in plastic.
  2. Plastic is not a conductor of electricity.
  3. Light is generated by a current of electricity passing through a coil in a bulb.
- Therefore;
4. No electricity will pass through the coil.
  5. The bulb will not produce light.

The other model introduced by Hempel, is similar to the above, but is inductive in approach, employing a statistical notion of probability to explain occurrences. Here instead of Universal laws as utilised in the D-N model, the degree of likelihood is cited to join the premises together, from which the explanandum may be inferred inductively. Needless to say, there is much discussion concerning the defects of these models. At best, these and other suggested models capture something like the shadowy intuition we have concerning the nature of explanation: there is as yet no completely satisfactory model of explanation with which one could conform<sup>4</sup>. However, there does appear to be a consensus that 'explanation' as an activity, does depend upon there being natural relations between events; where these relations are law like, or probable to some degree, in virtue of some properties of the objects cited in the event description.

It should be noted that not all explanations are causal explanations<sup>5</sup>; for instance, the explanation of why the sum of the

---

<sup>3</sup> Hempel 1962.

<sup>4</sup> This point becomes relevant in discussion of Epiphenomenalism in Chapter 7, Pp 188-190.

<sup>5</sup> See Lewis 1992b for a discussion concerning the truth of this claim.

internal angles of a triangle are one hundred and eighty degrees, is 'geometric' rather than 'causal' - nothing 'causes' the angles to comply with such a rule<sup>6</sup>. However, this topic concerns the use of causal explanations, particularly in relation to providing an understanding of how the mind affects the body. After all, the purpose of an explanation is to further our understanding of the particular phenomena being explained. So whilst we may have been provided with a satisfactory *explanation* of why nothing travels faster than the speed of light, it will only increase our *understanding* if we are already familiar with a certain amount of physical theory. What constitutes a good explanation can only be determined by the context in which that explanation is sought. Accordingly, the terms used in that explanation, may differ in sense whilst referring to the same objects, depending upon what motivated the requirement for the explanation.

This is an important consideration in the light of the very different responses we have to information about brain states and about mental states: because our understanding of the former is in terms that do not provide comprehension when applied to the latter, there is a temptation to conclude that the two states are utterly different. However, as argued by Hilary Putnam, how you go about providing a particular explanation, depends more upon the questions being asked than the facts available with which to answer them.

He takes as an example two different ways of explaining why a peg passes through one hole but not another. There are two ways of describing the situation; either according to the material structure of the constituents, or according to their higher-level geometric properties. Both, he admits, provide explanations; but they are different insofar as there are geometric laws revealed at the higher level, which would be concealed at the basic level. The ability of the higher level explanation to uncover these geometric truths, makes it a more valuable form of explanation. The basic level simply provides a terrible explanation. The more general the explanation, the better it is in terms of facilitating scientific enquiry:

The fact is that we are much more interested in generalizing to other structures which are rigid and have various geometric relations, than we are in generalizing to *the next peg that has exactly this molecular structure*, for the very good reason that there is not going to *be* a next peg that has exactly this molecular structure. So in terms of real life disciplines, real life ways of slicing up scientific problems, the higher level explanation is far more general, which is why it is *explanatory*.<sup>7</sup>

### Subsection 2 : Levine on the rôle of properties in causal explanations.

It is Levine's opinion that causal explanations are justified upon the basis of citing properties that relate to each other in

---

<sup>6</sup> For further discussion, see Ruben 1990, Chapter 7.

<sup>7</sup> Putnam. 1973. P 297.

consistent ways. He illustrates this idea with an example involving hydrogen and oxygen molecules: granted the knowledge we have of the relations between sub-atomic particles, and their propensity to acquire or lose electrons in the process of stabilising their electron shells, it makes sense that they should combine to form  $H_2O$ . In addition, it makes sense that  $H_2O$  should have the surface features that it has, such as its lacking taste, smell and being colourless, for instance. As we come to know more about the behaviour of very small particles, and the principles of combination - such as those concerning the filling of electron shells - so our ability to explain the grosser features of substances is increased. We progress from a system of pre-theoretic concepts of objects, based upon the way they manifest themselves to us, toward a scientific understanding of those objects. In this process of reduction, Levine claims that the concept to be reduced - such as 'water' - is seen as delimiting the range of causal relations in which that object enters. By investigating the sorts of physical structures that would explain the causal behaviour, we acquire a refined concept of that object. The relation between causation and explanation is neatly summed up as follows:

Our concepts of substances and properties like water and liquidity can be thought of as representations of nodes in a network of causal relations, each node itself capable of further reduction to yet another network, until we get down to the fundamental causal determinants of nature.<sup>8</sup>

### Subsection 3 : Implications for a Materialist analysis of mental causation.

In the light of the above, it is worth specifying the goal of Materialism in terms of how the properties of physical and mental states are held to be related. Consider two events, A and B. Normally we might suppose that there are some properties of the antecedent event, A, which in conjunction with the background conditions, cause the consequent event, B. If you combine two sets of properties, you get a reaction between them that brings about the consequent event. In explaining this relation between cause and effect, it is usual to refer to the properties and law-like theories that relate the two. When it comes to causal relations between the mind and the body, the situation is far more complicated. Even a cursory consideration of the terms in which the mental properties of sensation and cognition are described - predicates such as 'searing', 'hopeful', and 'misguided', for instance - is sufficient to demonstrate how absurd it would be to apply these to physical objects. Such is the divide between them, it makes the chances of coherence between the two improbable.

Therefore, when it is stated that a physical state has been caused by a mental one, or when a mental state has been caused by a physical one, the means by which this cause operates is unclear. Three possibilities arise: that the term 'cause' is being used

---

<sup>8</sup> Levine. 1993. P 132.



illegitimately, and the relation between the mind and the body is one of accidental correlation: that there is a causal relation, but it takes place between two distinct types of substance: that there is a causal relation between events of one type of substance, though it is by no means clear how the properties relate.

The Materialist denies the first and second possibilities and asserts the third. Importantly, they hold that the mental event caused has a physical description. The danger of this approach, one which Davidson is accused of ignoring, is that seems to follow that there is no justification for thinking that mental properties have any influence at all. What the Materialist requires, is that mental and physical properties are co-instantiated by a single event; that we can refer to that event either by its mental or its physical description; and that the mental properties we cite are efficacious rather than epiphenomenal.<sup>9</sup> The Materialist needs an account of how a sequence of physical events could lead to a mental event, which may be identified with the occurrence of a physical event. This identification is something that Kripke has urged is beyond possibility: in which case, so is the Materialist project. Therefore it is essential that the arguments of Kripke be re-examined.

## Section Two

The points of Kripke's argument in 'Naming and Necessity' that are of immediate concern, are those he makes about the problems facing Identity theorists. These remarks may be extended to Functionalist theories also, insofar as they identify such mental states as 'sensation' for example, with some physical process. Kripke's line of thought is concerned with the processes of scientific discovery, and as such, is concerned with epistemological matters. The argument Kripke presents, is taken as providing a strong case for denying the identity between brain states and mental states. The following considerations suggest that his views sanction only a weaker claim; that, necessarily, we are not in a position to make such an identification. If the latter can be demonstrated to be the case, then it is a first step towards a principled basis for embracing Agnosticism.

The remainder of Section Two will consist of a re-statement of Kripke's position, followed by an attempt to demonstrate that the scope of Kripke's objection is broader than he anticipated. As such, not only is it anti-Materialist, but it also implies that Dualists will have analogous difficulties defending interaction between the mind and the body. If we take his remarks to invalidate Materialism, we must consider what positions remain available to us. In order for the process of scientific discovery to occur, there has to be a strong relation between mental and physical states. If this relation is not one of identity, then perhaps it is a causal relation, as suggested by Substance Dualists. However, any such causal relation requires the identification of a physical property, which is not only sensitive to mental influences, but also capable of influencing mental states.

---

<sup>9</sup> A solution to this problem will be offered in Chapter Seven, Pp 186-190.

Arguably, their chances of identifying this property with any part of body, are as remote as those of the Materialist.

If we accept the logic of his argument, then there is no reason to accept an identity between a brain state and a mental state; further, there is no reason to suppose there is any commerce between these states; and more radically, one has no reason to accept any theories concerning the nature of the external world as we naïvely conceive it to be. Rather than adopt such a sceptical position, I chose to consider a different interpretation to the issues Kripke raises. Our inability to establish an identity, may be because no such identity exists, or because - for whatever reason - we are not in a position to make that identity. It is this latter possibility that will be examined in Section Three.

Subsection 1 : Kripke's remarks upon identifying mental with physical states.

Kripke acknowledges that identity theorists in the Philosophy of Mind, are after some form of identity analogous to other identifications that have been made by science. It is important that these identities result from a recognisably scientific activity. The example I shall concentrate upon, is that of the identification of heat with molecular motion. In order to distinguish between heat, and the sensation of heat, I shall adopt the practice of indicating the sensation with a subscript. Thus 'Heat' will refer to the phenomenon of heat, and 'Heats<sub>s</sub>' will refer to the sensation of heat. It is clear that Kripke regards these two as independent:

When I refer to heat, I refer not to an internal sensation that someone may have, but to an external phenomenon which we perceive through the sense of feeling; it produces in us a characteristic sensation which we call the sensation of heat. Heat *is* the motion of molecules.<sup>10</sup>

There are two reasons for accepting that Heat does not mean Heats<sub>s</sub>. Firstly, there might have been a world devoid of sentient creatures; provided that world contained molecules, then there would be some degree of heat within that world. So you would have Heat in the absence of Heats<sub>s</sub>. Secondly, Kripke conjectures that there could be a world where Heats<sub>s</sub> was experienced as a result of some change in the creature's environment other than change in molecular motion. In which case there would be Heats<sub>s</sub> in the absence of Heat.

The identification of Heat with molecular motion is the result of first detecting the presence of Heat through its property of causing Heats<sub>s</sub>, and noting that changes in sensation occurred correlatively with increase in molecular activity. The apparent contingency of the identification lies with Heat having the property that causes this sensation. It is this feature of the relation between Heat and molecular motion which is contingent.

---

<sup>10</sup> Kripke. 1980. P 129.

Is it reasonable to suppose that the suggested identity between Pain and C-fibre activity may be considered as a similar instance of scientific discovery? Kripke says not, for the following consequence of his theory of identity:

This means that the identity theorist is committed to the view that there could not be a C-fiber stimulation which was not a pain nor a pain which was not a C-fiber stimulation. These consequences are certainly surprising and counterintuitive.<sup>11</sup>

The contingency of the identity between Heat and molecular motion was shown to be a result of confusing the two types of heat - Heat and Heat<sub>s</sub>. Certainly, the identity between Pain and C-fibres appears to be contingent. However, Kripke notes that the same move cannot be made on behalf of this identity claim, as was made on behalf of the identification of Heat and molecular motion. There is a property of Heat to which we humans are sensitive. However there is no analogous distinction between Pain and a property of pain that we sense. Nor is there any sense to be made of the proposition that P<sub>s</sub>, qua sensation, is a contingent property of Pain. The idea of disembodied pains is absurd, so is the idea that something could have the appearance of pain without it actually being painful. If there is a sense of contingency in the identification of Pain and C-fibre activity, it is because there is no real identity between the two.

### Subsection 2 : Implications of the above for the viability of Science.

Three assumptions are made within the context of identifying Heat with molecular motion, which will not be contested. Firstly, in order to fix the reference of Heat by Heat<sub>s</sub>, we have to assume that there is a physical causal process that reliably leads from Heat to nerves, to central processing, to sensation. We must also assume that individuals do correctly and consistently identify the sensation Heat<sub>s</sub>. Finally, we must assume that the properties of physical substances behave in a nomologically satisfactory manner, and that there is a significant degree of regularity according to which their properties might be incorporated into a theory. These assumptions are required if one is to entertain the possibility of making scientific discoveries and devising theories. However, it is the object of what follows to tease out tensions within Kripke's argument. In effect, the first and last assumptions are at odds; put boldly, the scientific discovery above involves a non-contingent relationship between Heat and Heat<sub>s</sub>, but Kripke's argument generalises to deny that such a relationship exists.

Molecules are taken to have a number of properties, some essential, some contingent: for instance, it is taken to be an essential property of water molecules, that they are combinations of hydrogen and oxygen; alternatively, it is a contingent property of

---

<sup>11</sup> Kripke. 1980. P 149.

these molecules that they are frequently used as examples in philosophic discussions. Consider the relationship between Heat and molecular motion. According to Kripke, Heat *is* molecular motion. The sensation Heat<sub>s</sub>, that we refer to in English as 'Heat', is related to the phenomenon of molecular motion in such a way that this sensation changes as the rate of motion changes. Heat exists where molecules exist. It will affect other molecules in such ways as to cause 'expansion', 'contraction', 'evaporation' and so forth. In principle, a thorough scientific understanding of molecules will cover the effects Heat has upon the molecules that make up our nervous system. Changes in Heat will therefore be reflected by changes in our nervous system. Let Heat<sub>B</sub> refer to the particular brain state, associated with Heat<sub>s</sub>, that is affected by Heat. According to our theory of molecular behaviour, there is a nomological link between Heat and Heat<sub>B</sub>. According to Kripke, the relationship between Heat<sub>B</sub> and Heat<sub>s</sub> is merely contingent. That is to say, we can imagine creatures who experience Heat<sub>s</sub> in circumstances where Heat<sub>B</sub> does not obtain, as well as those who have no experience of Heat<sub>s</sub> when their brains were in state Heat<sub>B</sub>.

However, it is an assumption that our experiences may be taken as reliable indicators of the way properties behave. Changes in Heat<sub>s</sub> correspond to changes in molecular motion. This is required if there is to be such a thing as empirical study. Without the faith in there being a law-like relation between our sensations - be they perceptions or feelings - and the physical world, then our theories concerning the nature of that world fall short of the rigour to which we aspire with science. Our theories concerning the nature of Heat, depend upon our being certain that what we are investigating in our experiments is actually Heat, and not some other phenomenon. If we cannot depend upon Heat<sub>s</sub> actually picking out Heat as a phenomenon, then the results of our studies are valueless. In which case, we either must treat the relation between Heat and Heat<sub>s</sub> as being something more than merely contingent, or else accept that the pursuit of knowledge through scientific activity is a waste of time.

### Subsection 3 : Implications for the viability of Dualism.

We might ward off the attack on science by admitting that there is a link between Heat<sub>B</sub> and Heat<sub>s</sub> that is stronger than contingency, but weaker than identity. One might suggest that there is a causal relation between the two. This is less than the Materialist wants, but is perfectly satisfactory so far as the Dualist is concerned. The following is an attempt to demonstrate that Kripke's point has some force against this response, concluding that no matter which position you take, there will be problems concerning the relation between physical and mental states.

Both Materialists and Dualists agree that there is interaction between the mind and the body. It is also important that the mind is not epiphenomenal with respect to what happens at a physical level: the mind must 'make a difference'. Both positions cannot allow that

there is simply a brute correlation between mind and body, without conceding that we are incapable of deliberate action. If we are forced to make that concession, then we must accept that the pursuit of science is wholly beyond us, since the pursuit of knowledge through science requires the exercise of reason. In the above discussion, the physical state  $\text{Heat}_B$  interacts with  $\text{Heat}_S$ . For the Materialist,  $\text{Heat}_B$  simply *is*  $\text{Heat}_S$ . Dualists hold that  $\text{Heat}_B$  causes  $\text{Heat}_S$ . If Dualists want to retain a non-contingent link between mind and body, then this has to be reflected in the link between, for instance,  $\text{Heat}_B$  and  $\text{Heat}_S$ . In which case, it cannot be a contingent property of  $\text{Heat}_B$  that it causes  $\text{Heat}_S$ .

One of the chief obstacles to an interactionist theory, has been the need to observe the principle of the conservation of energy. Recently, Lowe has offered an account which he claims conforms to the above principle<sup>12</sup>. Consider a movement of the arm towards an object 'x', that has been prompted by a desire. This may be described as a physical event 'E' and a putative mental cause 'M'. He notes that were we to trace the cause of 'E' through the maze of prior neurophysical states, we would end up with an unconvincing disjunction of terms. This contrasts with the ease with which we can identify a cause in mental terms; simply, that 'E' is caused by a desire for 'x'. On this account, the events in the brain are organised in such a way as to prompt 'E'. The events prior to 'E' are inter-linked in such a way as to comply with the principle of conservation of energy: that is, the influence of the mind does not interfere with the causal relations between physical states. Instead:

[M]...is causally responsible for the fact that *there exists a maze at all with that particular convergence characteristic.*<sup>13</sup>

The advantage of retaining such concepts as 'desire', 'belief' and so forth, is that otherwise, the patterns of processing in the brain would merely appear as formless activity; these concepts are needed to provide a classificatory structure, within which the brain processes become intelligible. Even so, there has to be a property of physical matter that explains how it is able to converge according to the influence of mental states. Therefore, it seems generally, that the Dualist is committed to there being a physical property that allows this interaction - call it  $I_P$ . In every nomologically possible world like ours,  $I_P$  has this property of enabling the link between the mind and the body.

The problem is as follows. The identity between C-fibre activity with pain was shown to be problematic, because we can imagine cases where pain was present and C-fibres inactive, and pain absent when C-fibres were active. However the possibility of identifying any physical property with  $I_P$  appears similarly bleak; it seems that we can imagine mind-body interaction without this property, and this property where interaction is absent. Whilst it is not strictly the Kripke identification of a mental state with a

<sup>12</sup> Lowe. 1992, 1993, and 1996 Chapter 3, Pp 82-90. Another example may be found in Hart 1988.

<sup>13</sup> Lowe. 1992. P 274. This position is criticised below; Chapter 7, Pp 195-196.

physical one, it is none the less an identification between something picked out by physical vocabulary, with something that is picked out by mental vocabulary - that is, as a physical property that is sensitive to the influences of the mind. According to Kripke, for any identity relation we make, the objects thus identified are necessarily identical.

Unlike the Materialist, it may be granted that the Dualist is not saddled with having to identify a physical state with a mental state: but it seems that their problem is having to identify a physical state with a certain property that allows for the interaction between the mind and the body. For any physical state they come up with, it is possible to imagine that either interaction can take place in its absence, or that interaction need not occur in its presence. It does not appear possible that we could identify a physical property that is capable of affecting - and being affected by - mental properties. Hence, belief in interaction is unwarranted, in just the same way as belief in mind-brain identity. Therefore even Dualists are faced with the difficult and counterintuitive problem that Kripke has set Materialists.

#### Subsection 4 : Is Science compatible with a contingent mind-body relation?

The problem that the Kripkean faces, is that in order to have a scientific practice which endorses identity statements such as 'Water is  $H_2O$ ' and 'Heat is Molecular motion', they require a suitable link between the sensation and the presence of the object sensed. If they allow there to be such a link, then they are committed to identifying a connection between the physical and the mental. If they deny the existence of the link on the basis of their arguments concerning necessity, then they must abandon their faith in scientific activity as a means for discovering truths about ourselves and our environment. If we wish to retain our faith, then we must find some way of explaining the apparent contingency of the identity between some physical states and mental states.

So, a great deal depends upon how the notion of 'contingency' between the sensation and the object sensed is being construed. Kripke's argument points to three ways that the relation strikes us as contingent. Firstly, we realise that the presence of molecular motion is entirely independent of its being sensed. Secondly, it strikes us as reasonable to suppose that the presence of Heat as an external phenomenon might be associated with some other sort of sensation by a different creature from ourselves. Thirdly, causing the sensation  $Heats_s$  seems to be an accidental property of Heat; it might have been the case that humans were sensitive to light according to presence or absence of  $Heats_s$ . The question is, how fatal are these senses of contingency when considering the relation between Pain and C-fibre activity?

Clearly, for the Materialist,  $Heats_s$  will have to be identified with some type of brain activity. Again, let us suppose that  $Heat_B$  is the suggested brain state to be identified with  $Heats_s$ . It is argued

above that there will be a theory which accounts for the behaviour of molecules, explaining how they affect the cells of the skin, resulting in changes within the central nervous system. The theory will refer to properties of the various objects involved in this sequence of events. The relation between these properties is not contingent, but nomologically necessary: in the same way, properties of hydrogen and oxygen, as employed in explaining their behaviour, are taken to be non-contingent. In which case, the first construal of contingency applies only to the coincidence of heat and creatures who are sensitive to heat. What it does not imply is that where you have heat and organisations of matter similar to human organisms, it is only contingent that their bodies will behave at the physical level the way they do.

The second application of contingency is similar to the first, in that it refers to the case that there are other forms of action between molecules, which result in different systems reacting in systematic ways with the variation of molecular movement. Again this does not affect the conditional assertion that the human system will react thus and so in respect to the movement of molecules.

It is the third application of contingency which is most challenging to the Materialist. For it suggests the possibility that for two physically identical creatures, within the same nomological circumstances, it is contingent that one creature should have one sensation whereas the other should have either a different sensation or none at all. (These possibilities are exemplified in the problems of inverted qualia, and Zombie-hood, respectively). It is a consequence of this possibility that our faith in science becomes undermined, for it is the sensations that provide the initial evidence of some external phenomenon. If the sensation is linked only contingently with some physical state of the organism, then there seems to be no justification for any identifications made by the senses between surface features and their underlying physical states: if the same sensation type could be associated with different physical states, then there would be no reason to identify it with one state rather than another. The implication of Kripke's argument is that if we are to doubt the nomological dependence between  $\text{Heat}_B$  and  $\text{Heat}_S$ , then we should also doubt such identities as 'heat and molecular motion', and 'water and  $\text{H}_2\text{O}$ '. This is not to say that Kripke is wrong, but that if he is correct, then there is a lot more at stake than simply the Identity thesis: all scientific endeavour is at a standstill.

#### Subsection 5 : Conclusion.

If we ignore the quality of the sensation for the moment, it appears that the relation between the properties of the physical constituents must behave in a law-like manner if science is to be a viable activity. Furthermore, sensations have to be reliable and consistent indicators of phenomena if scientific practice is to be justified. The identities cited as examples of scientific discovery (Such as: "Heat is molecular motion".), depend initially upon

instances of sensations, and the phenomena of which those sensations are appearances. If the identity of the sensation with the underlying physical stuff is questioned, then so too is the nature of the reality of which the sensation is an appearance. This is a consequence that follows regardless of whether one holds to a Materialist or Dualist position. If such an identification cannot be made from any available stance, then it is worthwhile pursuing the following idea: rather than conclude that there is no identity relation to be made, conclude instead, that we are constituted in such a way as to be unable to warrant believing in the identity relation.

### Section Three

Here, the argument engages with remarks that Nagel makes in reference to 'Appearance' and 'Reality', and the activity of science. Sensations are taken to be the appearances of a distinct Reality. We believe that Reality is different from how it appears to us, and therefore we need science to bring us toward a proper appreciation of it, by resolving the anomalies that contaminate appearances. Nagel supposes that since experiences are an incomplete representation of reality, they cannot be a candidate for scientific investigation; as such, since everything physical is knowable by some science, it follows that the physical nature of experience is beyond our comprehension.

However, sensations are not appearances, as Kripke notes. Two things can appear alike yet be different - such as 'gold' and 'fools gold'. However, anything that resembles 'pain' simply *is* 'pain'. It should be noted that Rosenthal<sup>14</sup> suggests that there is a sense in which we use the word 'Pain' on occasions when we are unaware of any such sensation; for example, when concentrating intently upon some book, we will not always notice how uncomfortable our posture feels. This use of pain, he notes is not identical with Kripke's use. However, since we are presently concerned with the relation between consciousness and physical states, this observation is of little comfort to the Materialist. Anything resembling our experience of pain will be painful.

#### Subsection 1 : The rôle of sensations in scientific progress.

Sensations are not appearances, but are the effects of physical objects upon our senses. The appearance of the stick that is half submerged in the water, is commonly offered as an example of why we should not trust our senses. Instead of appearing straight, it appears bent just at that point where it enters the water. The following thought motivates a distancing of our theories from our perceptions. If we are to depend upon our senses, then we may assume that water has the property of so acting upon solid objects that it distorts their shape. There is a possible world where doubtless there is such a liquid substance which has that effect on

---

<sup>14</sup> Rosenthal. 1994. P 354.



solids, but we would want to say that it is not a property in our world that such an occurrence is nomologically possible.

In the above example, we want to say that there is no bent stick. In reality, the light waves reflected from the stick do not enter our eyes in a uniform manner. Hence, because water refracts light differently from air, there is a change in perspective as the reflection passes through the distinct media of air and water. The appearance is not at fault: it is a consequence of such things as the properties of light, water, and our visual system, for instance. The error lies in the inference we draw from the perception, when we conclude that 'rigid objects do odd things in liquids'.

The perception of the shape of the stick suggests one of two things: liquids do strange things to solid objects: liquids do strange things to our ways of seeing things. In both cases, the appearance provides evidence that may lead to an important scientific discovery. Both hypotheses are supported by the appearance; that is, they explain how the stick appears the way it does. In order to provide evidence for these hypotheses, the appearance must be taken as a consequence of the conjunction of the stick and water. To that extent, the sight of the bent stick cannot be treated as an illusion, if by illusion is meant 'deceptive appearance' or 'false perception'. If we take the absurdity of the first hypothesis to show that our senses cannot be trusted, then the second hypothesis cannot take the way the stick appears to us, as trustworthy data either: and this disqualification of evidence, would have the unsatisfactory result of forcing us to abandon some of our theories of optics. Scientific advancement requires that we take the evidence of our senses at face value, and become more wary of the inferences we are inclined to draw as a result of those experiences. In which case, 'experience' is as objective as it needs to be, and should not for that reason at any rate, be barred from being used as evidence in scientific investigations.

### Subsection 2 : The puzzle of understanding the nature of sensations.

The difficulty we encounter when attempting to provide a scientific understanding of sensations, does not lie with the objectification programme suggested by Nagel. Rather, it is simply that the procedure of objectification is not something which can be executed with respect to our sensations. The fact that science will fail to uncover the nature of qualia, need not show that qualia are not physical in nature, unless one can argue that anything physical can be investigated successfully by science. For it could be that qualia stand in the wrong relation to our intelligence to be investigated, and therefore - in scientific terms - they are unknowable. What eludes our understanding, is the nature of the properties that tissue has which are responsible for generating sensations. This is not because these properties are not physical; as was pointed out above, on any account of the Mind-Body relation, there has to be some physical property involved in the link between

the physical and mental. Rather, our ignorance stems from the consequence of this property not being sensible property.

In order to get some intuitive grasp of this suggestion, the following is a naïve account of what is meant. Our understanding of the world is limited by those aspects of it to which we are sensitive, either directly or indirectly. What happens when we turn our attention to the process of sensations, is that we either get hold of the bits we can see or detect, or we can make up theories to explain the observed links. This is suggested by the image of the torch in the attic, which acquaints us with objects by picking them out with its beam. What we think is that the torch is part of the contents of the attic, and so there must be a way of becoming acquainted with it also. This would mean turning the beam upon itself. Even if you could make sense of being able to do that, there is a further difficulty. It is assumed that as the beam is of the same nature as the contents of the attic, it is therefore susceptible to scrutiny by a beam of light. Nevertheless, you would not necessarily have a clear understanding of the nature of light by shining another torch at it, insofar as what needs to be understood, is the relationship between the beam itself, and the effect it has upon the objects it falls upon.

In some ways, this is perhaps the root of the problem; that we expect to elucidate the nature of experience - with an assumption that experience is part of the natural world - in the same way as we are acquainted with the rest of the world, which is to say, via our senses. We have no organ that can discriminate qualia, therefore we cannot investigate them in order to determine their structure: it is like being asked to 'sense' the 'sensing' of an object.

Suppose the following is a 'folk' description of the strategy we employ in the course of understanding phenomena. We are sensitive to various properties of objects; that is to say, we have such sensations as sights, sounds, smells, tastes, and feels. The properties to which we are sensitive, are individuated by recognitional concepts - such as 'this feeling belongs in the same category as 'that' sort of feeling' - and full-blown concepts - such as 'that is the scent of chrysanthemums'. These develop into hypotheses and explanations - such as 'water becomes ice at a certain temperature, because at that temperature the molecules have insufficient energy to break the lattice-like bonds between them'. Patterns emerge, and derived from these, laws governing the behaviour of properties. In this manner, it is possible to identify such things as 'water' and 'H<sub>2</sub>O'. The 'water' concept refers to the surface features to which we are sensitive - such as its colour, taste and touch at various temperatures. The web-like structure of our scientific beliefs, places these properties within a chemical context, such that we can identify a particular compound with the substance that behaves in the way that 'water' does; because Hydrogen and Oxygen have the properties that they have, it is obvious why 'water' has the properties it has.

So if we pursue this strategy with respect to an investigation into the nature of 'experience', then the analogous first step is to enquire how do 'experiences' affect our senses? But this question is

absurd; experiences just are the affecting of our senses. Consider again the identification of water and  $H_2O$ . This may be thought of as an equation of properties. On the one hand, the properties of Hydrogen and Oxygen (considered as theoretic entities) are such that in certain specific circumstances, they will behave in predictable ways - for instance, they combine with each other, evaporate, flow, and do not stimulate the taste buds in certain living systems. Water, as picked out as an perceptual concept, has such properties as being tasteless, solid when it is cold, liquid when it is warm, and steam when it is very hot. Hence there is a match between the two ways of picking out water by theory and experience; the correctness of this match is strengthened by its supporting predictions, and its inclusion in the explanations of other phenomena.

What we lack are suitable equivalencies between experience and physical stuff, in an analogous way. The above strategy does not work when applied to 'experience': we have no idea which properties of physical matter are capable of being the various types of sensation available to us. This need not be taken to show that mental states are not physical; it could be that they do not occupy the appropriate position in the explanatory scheme. In short, we need to know the effects mental states have on us in order to match up the properties. 'Experiences' do not have effects in this sense, so the investigation never gets off the ground.

#### Section Four

It remains to show how the above conclusion relates to the uneasiness we might feel when thinking of Mental states as being realised by physical processes. In particular, there is bound to be a residue of disbelief in the ultimately physical nature of intentionality, of sensation and thought, as well as the irreducible indexicality of the self. In what follows, I want to sketch an interesting relationship between these three properties, and the process of understanding.

Some philosophers have suggested that the functional aspect of consciousness can be replicated by mechanical processes. As a consequence, there is at least a theoretical understanding of how 'learning' and other intelligent processes can take place in physical terms<sup>15</sup>. What has frustrated a complete physical understanding of the mind, is the apparent impossibility of understanding how 'consciousness' could be a product of activity in the brain.

##### Subsection 1 : An analysis of theory formation.

It is arguable that three elements are required for the formulation of theories about phenomena; firstly, that there are systems providing information about the phenomena; secondly, that there are systems capable of ordering this information, and using it in inferences to produce further information; and thirdly, that these systems are co-ordinated to form some sort of unity. These three

---

<sup>15</sup> See especially Fodor and the 'Language of thought' hypothesis - Fodor 1975.

elements appear in machines, particularly computers; and in humans, it is suggested that these elements are realised by the properties of 'sensation', 'thought' and 'personhood'. In addition, one must assume that there is some co-operation from nature, insofar as there are law-like regularities to be understood by the above types of system.

The three factors are jointly necessary to the ability to take stimuli as input and convert it into behavioural output according to some goal of the entity. For if stimuli are absent, there will be nothing for the inference generators to process. If the property of inferring is absent then there will be no means of deriving information from the stimuli - it will be simply a bland 'noise'. If the sense of unity is absent, then there is no grounds for treating the stimuli as belonging together in any useful sense, to allow any inferences to be made. These remarks apply as much to intelligent systems as to animals.

It is important to distinguish between the propriety of the application of the term 'intentional', and the ability for physical systems to realise intentional systems. It is rightly objected that the ascription of 'intentional properties' to machines, is a result of a human interpretation of their behaviour. As such, the ascription is parasitic upon the notion of human intentionality. Consequently, artificial intelligence cannot be regarded as a straightforward analysis of intentionality, without leading to the following absurdity:

If an assembly of physical elements is to be seen as a collection of mental states, it must be given an interpretation. This, however, requires an act or source of interpretation which...must be mental if anything is. But in that case it has a reality only under some interpretation which, in turn, requires an act of interpretation, and so on *ad infinitum*.<sup>16</sup>

Whilst it is true that machines have no awareness of their intentionality, this does not mean that physical systems cannot act in an intentional manner. The point is, that it is possible to simulate intelligent tasks using physical mechanisms. That is, the resources of physical matter are rich enough to instantiate intelligent systems. If intentionality can be realised in computers, then why cannot intentionality be realised in a physical conception of humans?

A human is sensitive to a number of aspects of the world via the senses. These aspects are the basic phenomena to be explained. The information provided by the senses alone would not suffice to formulate theories, since there needs to be a process whereby the information is divided up into elements, which serves as a basis for classification into groups. Once classified, it is then possible to mark patterns in the way these elements occur, leading to the formulation of laws. There also has to be a way of drawing inferences from general laws to particular instances; both these processes may be accomplished by our ability to think. It is a presumption that all thought is conscious, but this has been seen to

---

<sup>16</sup> Lund. 1994. P 123.

be erroneous<sup>17</sup>. It is possible to behave in a manner that evinces deliberative planning, without the agent having been aware of such preparation; the driver example of Armstrong's provides an instance of this.

So far as the intentionality and logical nature of thought are concerned, whilst there is nothing in physics that corresponds with these defining properties of thought, it would be an error to confuse 'thought' as an object, and 'thought' as a process. Arguably these properties belong to the process of thinking, rather than to anything that might be individuated as 'a thought'. The term 'Validity' applies to a relation between propositions rather than to the processes by which these propositions come to be expressed. For instance, the premises and conclusion of an argument can still stand in a logically valid relation to each other, even though they are expressed in different languages. Certainly, there is nothing in the underlying properties of matter from which the brain is composed that is intrinsically logical. If there were, then perhaps our thinking would always be valid. The validity of conclusions is a property of the correct ordering of thoughts, rather than in a causal relation from one thought to another.

In order for thinking to operate on the sensory data, it is necessary that these experiences are unified; that what joins two experiences together is the fact that they were experienced by the same system within a temporal framework. For example, that we are able to form theories about the Doppler Effect, depends upon treating the various discrete sounds as being united rather than separate. Further, we have to be in a position to connect the colour of certain flowering plants with the scent of those plants, in order to form theories about nocturnal insect pollination. Unless it is possible to link the sensations within a sensory modality, as well as those between modalities, there would be no way of making inferences required in the formulation of theories. What makes a series of sensations an 'experience', is some means by which they are regarded as being part of a whole. From this arises our sense of individuality.

### Subsection 2 : A brief sketch concerning the nature of the emotions.

Parenthetically, drawing upon these considerations, one might justifiably entertain the following idea concerning the nature of the emotions. They are a complex sensitivity to events affecting one's self, arising from an interplay between the unity of the perceiver and the recognition of that unity's relationship with the world and other people in it. In this way, the emotions are sensitive to potential harm and benefit to the self, just as the body is sensitive to damage by feelings of pain. The feelings associated with the various emotions are the result of physiological changes, that prepare the individual to acquire the potential good or avoid the potential harm.

---

<sup>17</sup> See remarks by Rosenthal above; Chapter Three, P 68.

These ideas are explored further in "Emotion and Feeling", by G.Madell. In this article, he develops an analysis of 'desire' introduced by S.Schiffer, in "A Paradox of Desire". Schiffer makes the following distinction:

There are two kinds of desires - those which are self-justifying and self-referring in a certain way, and those which are neither self-justifying nor self-referring at all; I shall tendentiously refer to the first as "reason providing (r-p)" desires, and to the second as "reason-following (r-f)" desires.<sup>18</sup>

In the first case, the value is conferred upon an object 'O', simply in virtue of my desire for 'O'. Where my desire is absent, so too is any reason to value 'O'. The desire 'provides' the reason for pursuit of the object. In the second case, my desire for certain objects, or states of affairs - such as good health - 'follows' from there being good reasons for valuing these things. In this sense of desire, the object only has a value insofar as I can appreciate the reasons for wanting it. What sets these two sorts of desire apart, is that there is a sensuous quality to the former, which may be absent in the latter. The significance of this difference lies in the rôle the feeling plays in our actions. Schiffer identifies the feeling in r-p desires as being either the pleasure experienced upon the satisfaction of the desire, or as the relief from a privation that results from gaining the object of desire. Hence, the r-p desire to do X, is the same as the desire for the experience of pleasure, or the relief from discomfort.

Madell notes that this account of the rôle of feeling has problems. Our desire to perform some activity, or realise some state of affairs, seems to be distinct from a desire for a particular state of consciousness. Schiffer's analysis ignores the fact that we are not indifferent to the process by which an experience of pleasure results. Were the desire to do X identical with the desire for pleasure, then any other activity Y that produced pleasure, would satisfy that desire. This contradicts our belief that the way we gain pleasure is an important part of our lives; though the experience might be qualitatively identical, there is a difference between the elation caused by overcoming a difficult obstacle, and that caused by taking drugs. Rather, the feeling is a response to the presence of an object or state of affairs, that potentially contributes to our sense of well-being:

Feeling is an essential element of r-p desire, not because the necessary *object* of such desire is a state of feeling...but because r-p desire is a form of *feeling towards*. Such desires are states of consciousness which are indissolubly both intentional and affective.<sup>19</sup>

This idea of 'feeling towards' ties in with the idea of our emotions as being sensitivities to states of affairs that resonate with an individual's sense of security. Just as pains alert one to the

---

<sup>18</sup> Schiffer. 1976. P 197.

<sup>19</sup> Madell. 1997. Pp 152-3.

possibility of physical damage, and thereby give one a reason to investigate and remedy the cause of the pain, so in this sense, the desires or aversions alert one to the possibility of psychological damage. For instance, the feeling of jealousy provides a motive to murder one's rival. Similarly, we can choose to ignore the physical pain if we believe that the nature of the remedy outweighs the inconvenience of the damage; so too, we can choose to ignore the prompting of our emotions, where we believe that such actions would result in the conflict of further desires - that of staying out of prison, for instance.

From this point of view, there has to be a distinction between those things I value by virtue of my having a felt desire for them - 'first order' desires - and those things which I am able to argue for as being valuable in an objective sense, regardless of how I feel about them<sup>20</sup>. The desiring of the 'first order' desire, was a positive evaluation of one of a particular range of desires, rather than a felt attraction to one of them. The wanting of that particular desire to become one's effective will, was implicit in the judging of its being objectively desirable, as being the goal of the sort of character that I would like to be: for instance, my choosing to continue working when I could have stopped and done something more immediately appealing, because I want to think of myself as a conscientious person. I make this choice because I realise that such an ideal is not fostered by sloping-off when there is work to do.

### Subsection 3 : A difficulty for Materialist theories of emotion and a solution.

Critics of Materialism have observed that the situations to which we respond emotionally, have a classification that is couched in a mental vocabulary. So, for example, we are 'dismayed', where we find our expectations have been far greater than the situation in which we now find ourselves. If the mental is ultimately physical, then there ought to be a physical analysis of the above, in which the mental term 'expectation' has been replaced. However, there is no recognisable Natural kind term which identifies such states; there is nothing but an arbitrary disjunction of physical descriptions of occasions of 'dismay'. Therein lies the problem:

Since such a grouping is not intelligible from the point of view of physical science, we are apparently forced to the conclusion that what gives the grouping the unity which it nevertheless clearly has is just the fact that the various physical states of affairs described are all manifestations of, or otherwise linked to, one and the same *state of consciousness*.<sup>21</sup>

Thus the Materialist appears to be putting the cart before the horse, when it comes to the identification of emotional states with any concatenation of physical properties. Yet it may be that the above consideration has less appeal when considering less

---

<sup>20</sup> For a discussion on this distinction, see Frankfurt 1971.

<sup>21</sup> Madell. 1988. P 38.



sophisticated states. There are fairly conventional physical groupings, along biological lines, which refer to the relation between the environment and the creature's needs. Assuming plants have no states of consciousness for the moment, there is no difficulty in specifying what is good and what is bad for them in physical terms, which are untainted with covert appeal to conscious states of pleasure and pain. Certainly, animals have motives for avoiding or promoting such sensations, but they also have those motives with respect to the physical states which - not to beg the question - are associated with pleasure and pain sensations. Taking emotions to be sensitivities to the environment in a much more complex sense, it is not surprising that there is no clear Natural Kind associated with them. When we experience an emotion on the above account, we are being sensitive to a complex property of the environment in relation to our values - whether it is conducive to them, or hostile. The sensing of humiliation is analogous to the sensing of colours, sounds, surfaces, scents and tastes; and these latter senses arguably do correspond to physical kinds.

Although this is purely speculative, it is possible to interpret human 'values' as having a physical correlate in terms of 'resistance'; that is one could replicate human choices based upon value, with a machine whose procedures had weightings built in to them. In this way, it would be physically harder, though not impossible, for the machine to perform one function rather than another. Admittedly, it would only be making 'choices' in virtue of our interpreting its behaviour in this way - taking the Intentional stance. But again, however inappropriate the term 'choice', there is no doubt that machines can vary their behaviour according to their environment, where the environment is classified according to physical groupings with which the machine can operate.

On this account, if by 'the autonomy of the mental' is meant the irreducibility of sensation, thought and self, then it is clear that the reason for this lies not in their being ontologically distinct, but rather in their playing an essential rôle in the processes of empirical understanding. The reason why we are unable to get a satisfactory understanding of the mind in an analogous way to our understanding of the brain, is because the components of the mind are precisely those means by which we are capable of understanding anything. To render the physical operations of the mind lucid, would require an independent method of 'understanding' - and that does not exist outside our senses. This explains why we take the relation between mental states to be characteristically conceptual<sup>22</sup>. Because of the necessary ignorance of the nature of the mind, only a cognitive appreciation of mental processes is available to us.

#### Subsection 4 : Nagel and problems with a Materialist analysis of the 'self'.

This approach may satisfactorily address the problems raised by Thomas Nagel.

---

<sup>22</sup> See especially work by Davidson (Davidson 1980). Also Pettit 1986.



In "A View from Nowhere", he writes:

Given a complete description of the world from no particular point of view, including all the people in it, one of whom is Thomas Nagel, it seems on the one hand that something has been left out, something absolutely essential remains to be specified, namely which of them I am. But on the other hand there seems no room in the centreless world for such a further fact.<sup>23</sup>

Nagel's point is that for any physical description we care to offer, there is only a contingent link between that description and the individual who is Nagel. We think of ourselves as being unique, in a way that conglomerations of matter are not. One can imagine moments when the physical description of our world is the same as other possible worlds. However, on that basis alone, we would not want to identify ourselves with individuals at those worlds, if their development were different from ours, for example. This uniqueness can hardly be captured in physical terms, and so any physical description will be incomplete.

Insofar as there is no physical property essentially coupled to an individual, in the way that the configuration of sub-atomic particles is definitive of an element, then it is true that given a physical description of the world, one would not be able to individuate people as one could individuate elements. Yet his difficulty stems from two problems. The first is that we have yet to formulate a satisfactory metaphysical definition of selfhood, against which to assess the measure of success of any proposed account. The second is a tendency to identify the organic whole of the individual with their 'self'. On the account adopted in this section, 'selfhood' is not a property of persons, but rather a relation between various systems, which serves to unite them in some common form of existence.

There is some empirical support for thinking that the individual self is a useful fiction. There is evidence to suggest that an individual may adopt more than one persona. For example, Dennett<sup>24</sup> discusses two phenomena that suggest that the principle of allocating one self per body cannot always be applied. The first is Multiple Personality Disorder: "...in which a single human body seems to be shared by several selves, each, typically, with a proper name and an autobiography."<sup>25</sup> The second involves people who have undergone a commisurotomy in an attempt to relieve their epilepsy. In both cases, patients behave in ways that manifest a different character from normal. According to Dennett's picture of consciousness, there is a 'centre of narrative gravity' which becomes identified with the self. However, the process that constructs this centre, can under stress produce other centres as well. Dennett suggests that MPD occurs as a result of childhood trauma. Victims create new 'selves' in order to identify the horrors

---

<sup>23</sup> Nagel. 1986. Pp 54-55.

<sup>24</sup> Dennett. 1991. Pp 419-426.

<sup>25</sup> Dennett. 1991. P 419.

they endured, as having happened to a different 'person'. The implication of these cases, is that if the self is a construction, then the uniqueness of the self is going to be dependent upon the environment in which the individual developed; and for this, there surely will be a physical description.

On a homelier note, we are often subject to visions and sounds that are beyond our conscious control. Sometimes, just as I am falling asleep, I have the experience as of overhearing conversations. Occasionally I recognise the voices, but not always. The topics of conversation always come as a surprise to me, and are not within my conscious control; I cannot make the interlocutors talk about what I want to hear, for example. In general, this is somewhat like dreams, where action very often involves scenes and dialogue that the dreamer feels powerless to alter. Yet there is some process by which these experiences are produced. The dialogue I 'overhear' has all the hall-marks of intelligence - I can understand what the people are saying. It is reasonable to maintain that in one sense of 'I', I am not responsible for that dialogue: I am just a passive witness to it. Yet clearly, in another sense of 'I', I am responsible: the experience has to be caused by something, and the interlocutors do not exist. We cannot, therefore, group together all our conscious states as belonging to one individual. The above is an example of a conscious state - of 'overhearing' - that does not belong in that group. This strongly suggests that there is a dominant 'I', but it denies that there is a discrete 'self', for which there has to be an essence. Rather, it appears more likely that it is a feature of sophisticated systems, that their cognition depends upon the ability to unify their experiences and thoughts.

Moreover, the idea of an understanding from no particular point of view whatsoever, is incompatible with the proposition that some measure of unification - some individuating perspective - is a prerequisite of understanding anything at all. It is one of the attractions of some forms of Idealism, that they make the point that any description of the world must be from some point of view<sup>26</sup>. The goal of scientific objectivity requires that our theories should not be limited by any individual's view of the world. What is not so clear is that this means that theories can go beyond every creature's viewpoint, where a viewpoint is defined in terms of the properties to which that creature is sensitive. Since we can only theorise about those aspects of nature to which we are directly or indirectly sensitive, we should bear in mind that it is possible that there are other properties of objects which do not manifest themselves to us in any way. Therefore we may acknowledge that our understanding of the world is incomplete. Yet this is not the same as having a theory concerning those possible entities. If they exist, the only property we can be certain of them having, is that of not manifesting themselves to creatures with senses like ours. In which case, Nagel is pursuing a vision of scientific objectivity which is wholly beyond us. Rather than our scientific understanding having the consequence of obliterating essential facts about ourselves, the above contention

---

<sup>26</sup> See Sprigge 1983, especially his fourth argument for Panpsychism Pp 110-131.

is that there could be no such understanding where something analogous to selfhood was absent. The relevant point is, that the self be seen as feature of a process rather than a property of an object.

One might add, that just as it is impossible to give physical descriptions of sensory and cognitive mental states because these stand in the wrong relation to our means of understanding phenomena, so too would we expect identity - the third feature - to be inaccessible to physical description.

#### Subsection 5 : Conclusion.

What the above concerns share, is a presumption about realization. Asking how consciousness arises from grey matter might be like asking how a brief case arises from bits of leather, thread and metal. The constituents of the briefcase do not have an essential and defining briefcase-ness about them. What they do have, is the ability to be fashioned into an object which performs the defining function of a brief case. Similarly, there is nothing 'mental' in the material that forms the central nervous system, but this material can be put together in such a manner as to perform the defining functions of the mind. If there is any doubt about this, then there has to be a doubt about the ability of computers to exhibit intelligent behaviour. What is unique about the mind, is its manner of gaining information about its environment in the form of sensory experiences; our understanding of the generation of these through neural processing is essentially beyond us, insofar as they play an essential rôle in understanding, and are therefore logically precluded from our comprehension.

### Section Five

The conclusion of the previous chapter was that McGinn's position was not exclusively Materialist. One reason for this weakness, is surely because it employs a theory, 'cognitive closure'. This theory utilised assumptions about mental properties, to motivate an argument to show that these properties are physical. However plausible the idea of cognitive closure is, its plausibility lies in consideration of mental properties rather than the nature of physical ones. As a consequence, all the theory shows, is that the relation between the mind and the body is beyond our intelligence; what it does not show is that the mind and the body are composed of the same material.

#### Subsection 1 : Is Agnosticism amenable to a Dualist interpretation?

The reason that New Mysterianism could equally be employed by a Substance Dualist, is that 'cognitive closure' is very much a mental property rather than a material one. If there is to be a convincing account of why the Mind-Body relation is essentially

beyond our comprehension, which does not lend itself with equal felicity to the Dualist, then it has to pay more attention to the relation between recognisably physical features and our understanding. Rather than explain this mystery in terms of our cognitive abilities, Agnosticism supposes that the explanatory gap appears because of a natural consequence of information processing. As such, this consequence applies as much to machines as animals, and is thereby independent of any implications arising from the presence of consciousness. By making the account independent of consciousness in this manner, it is disqualified from being considered as a potentially Dualist interpretation. Because it is a principle that applies equally to machines, there is nothing about minds per se which explains the mystery. To employ this principle in a Dualist account, would require a further argument to show why this story would apply in cases of distinct ontological kinds. However, there is nothing here that is incompatible with the truth of Dualism. The limited purpose of Agnosticism is to provide a response to the 'knowledge' and 'conceivability' arguments, rather than disprove any other position.

Is it possible that Agnosticism is exclusively Materialist in spirit? The following consideration suggests that it is not a position that Dualists would find attractive. It asserts that the ignorance of the link is implicit in the relationship between the objects to which systems are sensitive, and the composition of those systems. There is nothing intrinsically mental about this relation; it is supposed to hold in principle for detection systems where consciousness is absent.

Consider a robot,  $R_1$ , that is designed to pick up empty drink cans. It acquires knowledge about its environment through various light sensors. So far as it is concerned, its surroundings are composed only of objects whose properties are detectable by these sensors. Information from the sensors is loaded into its memory; this, in turn, is accessed by other processes in the execution of various tasks. There is a causal link between the light in the environment, the memory state and the tasks upon which the robot engages. For instance, a can reflects light into the sensors; these are linked to memory registers so that a certain configuration of cells are activated; a program compares this pattern against stored can-shapes to determine whether the object is suitable for collecting; having made a match, the program initiates motor procedures that result in  $R_1$  collecting the can. The designer of  $R_1$  is aware that the electrical properties of the components are utilised in setting up these links.

Consider a second robot,  $R_2$ , designed to diagnose faults in  $R_1$ . It does this by using light sensors to detect states of  $R_1$ , and compares these images with those pre-programmed into its memory. It is able to tell whether  $R_1$  is moving, by identifying the robot's shape on an array at regular intervals, and comparing the locations over time; and it can tell whether the various components are functioning, by comparing the actual state of the component with a catalogue of possible states – so it could identify a damaged wire,

or a broken fuse, for instance. From what it knows and what it can detect,  $R_2$  can trace a malfunction in behaviour of  $R_1$  – such as being stationary to going around in circles – to fault in its mechanism – by finding a mismatch between actual and expected shape of components.

If robots were given to such preoccupations,  $R_2$  might wonder how something like goal-directed behaviour could possibly result from different configurations of shape and colour. Whilst it can appreciate that there is such a correlation, what it cannot do, is comprehend the nature of the link between the behaviour and the state of the component. The sensor utilises electricity to provide information about the distribution and intensity of light. It cannot provide similar information about electricity except by inference. That is, it would have to know the mechanism whereby light wave energy was converted into electrical energy, in order to tell what the amperage was from prevailing levels of light. However, the system only provides information about light, and not about the way light is converted into electricity. Scanning the mechanism will only elucidate its visual properties.

In order for the current to represent the level of light, there must be some properties of light which interact with those of electricity, such as to allow for the state of the former to affect proportionately the state of the latter. Some mechanism converts light energy into electricity, at a fixed rate of 'x' candelas equalling 'y' amperes. Thereby, the patterns of light in the robot's environment, are represented by patterns of charge in the robot's memory. How this conversion takes place cannot be comprehended by the system, for it only appreciates half the equation; that is, the amount of light entering the system. It cannot grasp the other half - the current that emerges from the converter - as its detectors are not directly sensitive to such phenomena. Thus there is a gap in  $R_2$ 's understanding between the light coming in, and the changes of movement that are output; a gap which its means of comprehension will never bridge. The case is analogous to our understanding. As a Materialist explanation of our ignorance, it has the advantage over McGinn's by being distasteful to Dualist sensibilities.

### Subsection 2 : Is Agnosticism amenable to an Idealist interpretation?

Although the above reasoning does not support a Dualist view of the Mind-Body relation, can the same be said of other Monistic theories? In particular, could an Idealist use the arguments presented above, in support of their position? Hopefully, the answer is 'No', for otherwise, Agnosticism would not be a distinctively Materialist thesis at all. The following is a brief sketch showing why an Idealist would be inclined to shun this approach.

For the purposes of brevity, it will be necessary to blur the distinctions that exist within the group of views that are recognisably Idealist. Instead, it must suffice to concentrate upon the main difference between Idealism and Materialism, and employ

that distinction in deciding upon the propriety or otherwise of using Agnosticism. The many varieties of Idealism share a commitment to a division between the mind and reality, where reality is in some way dependent upon, and fashioned by, mental states and processes. Where there is no consciousness, there is no reality. Materialists assume that there is a reality that is sensed by the nervous system. That if the world were unpopulated, there would still be a number of objects all of which had properties which would interact with our sensory properties and produce sensations.

There is clearly a link between Idealism and Materialism on the above account. Insofar as the features of reality to which we are sensitive are dependent upon the relationship between properties of objects and properties of our senses, then it is obvious that our appreciation of reality will be constrained by the physiology of our senses. What is not independent of our senses, are those aspects of reality to which we are sensitive. However, what is independent is the objects to which those properties belong. The fact that we would not have a perspective upon reality without our sensory acquaintance, does not entail that there would be no reality where there is no sensing. We are logically precluded from forming a conception of reality except in terms of our senses. Yet, this does not mean that, for example, there could not be a type of property which, within a nomological kind, had no relations with those properties that affected sensory systems. Idealism assumes that all physical properties interact so as to make some impact upon our senses. This arbitrarily denies the possibility of sets of physical properties that interact completely in isolation from those sets which manifest themselves to us.

It is essential to Idealism that there should not be any properties that are outwith the scope of sensation or thought. For if there were, then such properties would constitute a reality that was separate from our perceiving or thinking about it. However, the above considerations conclude that there are such properties, since these are the ones by which consciousness arises from grey matter, and they are necessarily at the wrong end of the sensing relation to be a proper object of our senses. Hence, I cannot see that an Idealist would want to adopt the above explanation as a support for their position.

### Subsection 3 : Conclusion.

Agnostic Materialism argues that the property by which consciousness is realised in physical matter, is essentially beyond our understanding. It draws this conclusion from consideration of the relation between the properties to which we are sensitive, and the manifestation of this sensitivity. Our senses necessarily cannot scrutinise the nature of this manifestation in an analogous way to other phenomena, because it stands at the wrong end of the object-sensation process by which we become familiar with our environment. This is a consequence for any sensing system, not just animal ones.

As such, the nature of qualia depends upon the nature of the realising brain structure. For two individuals with the same type of sensory system, the only way the qualia could be inverted without showing up in their anatomy, would be if there were some fairly anomalous relations between substances. We may find this plausible if we are prepared to be sceptical about nomological necessity, where inversion is only a possibility in those cases where anything is a possibility. Again, there is no argument against this form of scepticism, but it does not pose a *particular* threat to Materialism in the manner envisaged by its supporters.

The problem posed by the 'knowledge argument' is addressed by this position, simply by denying Jackson's premise that Mary was in possession of: "...*all* the physical information"<sup>27</sup>. This is false, since there are some things about sensations that we logically cannot know, and that is bound up with the relation that they have to appearance and reality.

However, what still requires defending, is the idea that sensations are causally relevant to our behaviour. For the thesis that experience is realized by brain states, is consistent with the thesis that experience has no rôle to play in the determination of our actions. It is to this defence that the next chapter is devoted.

---

<sup>27</sup> Jackson. 1982. P 130.

## Chapter Seven

Even were it the case that consciousness may be regarded as a physical property, there remains the problem of knowing in what ways, if any, this property has a causal influence. Even if we allow for the fact that mental states are physical states in some sense of the term, that does not in itself settle the persistent doubt that the mind may turn out to be wholly without influence with respect to the rest of the body. In particular, there is a fear that the mind is epiphenomenal with respect to the workings of the brain. Generally, Epiphenomenalism may be seen as a form of scepticism about agency. Consider a situation that would normally be accepted as exemplifying an interplay between the mind and the body, which results in action:

A student in the Haldane library finds that the text they are studying has ceased to make any sense at all, even though they have put their glasses on. Dismissing the possibility that this unusual occurrence might be a product of their stupidity, they consider what other reasons there might be. Eventually, they realise that it is getting slowly darker, and the print is getting harder to see. So they turn on the lights and find that it is again possible to make sense of what they are reading.

In order to have some idea about the relationship between the desire state 'wanting to study' and the behaviour 'getting up to switch on the lights', we may consider it worth asking: "How can consciousness affect the brain?" Yet, Epiphenomenalism denies that there is an answer to this question, and insists that there is nothing about consciousness that explains why people behave the way they do.

The purpose of this chapter is to analyse the reasons behind this scepticism, and in so doing distinguish between two ways in which consciousness might be thought of as wholly irrelevant to the processes that sustain animal life. In its strongest sense, Epiphenomenalism asserts that there is only an accidental relation between the mind and the body. A weaker sense of the term allows that the relation is asymmetrical; that brain states can affect mental states, but that there is no reciprocal effect between mental states and brain states. For brevity, the respective positions on Epiphenomenalism will in future be referred to as Strong<sub>E</sub> and Weak<sub>E</sub>.

In order to place these distinct forms of Epiphenomenalism within context, Section One will briefly consider a number of arguments to explain why such a view is the object of so much attention. Responses to these will be considered, and rejected, in Section Two. Section Three will deal with the possibility of Strong<sub>E</sub>. As used in this sense, there is no causal connection whatsoever between consciousness and the brain. The fact that there



is a correlation between the two is only a mighty coincidence: nothing that we could ever discover would provide grounds for thinking that the relation of cause and effect held between mental and physical events. An attempt is made to explain why it is that Strong<sub>E</sub> is so difficult to refute satisfactorily, and concludes that the problem is sustained by one way of viewing the relation between the mind and the body. So long as this view is held, it will inevitably provide ground upon which a sceptical stance may be taken. However, drawing upon the conclusions of the previous chapter, it is by no means inevitable that this is the only view available; other views of the mind-body relation dissipate the threat posed by Strong<sub>E</sub>.

Though we may allow for the possibility of a satisfactory account of this relation, it will still be vulnerable to a sceptical interpretation. Weak<sub>E</sub> acknowledges that there is a causal link between the mind and the body, but holds that it only runs in one direction; that the brain produces states of consciousness merely as a redundant commentary, or produces them as a waste product. Weak<sub>E</sub> reduces the significance of consciousness to the level of a person's shadow. The movements of the shadow are dependant upon the light, and the movements of the person. There is a correlation between the movement of the shadow and the behaviour of the person, for the former may be explained in part by the latter; but the person's behaviour is independent of the movements of the shadow, and the shadow has no life of its own. By analogy, the mind is the shadow of the brain.

The fourth Section of the chapter provides some reasons to believe that this is an unacceptable interpretation of the relationship between consciousness and behaviour, and in conjunction with the conclusions of the third Section, it is concluded that Epiphenomenalism in either sense is not a tenable point of view. It is a consequence of the above discussion that it is our notion of 'causation' which is partly responsible for making Epiphenomenalism seem plausible. There are, however, other ways of thinking about causation; accordingly, Section Five briefly discusses the idea that consciousness might be an emergent property of the brain, before going on to consider Dretske's distinction between 'structuring' and 'triggering' causes.

In a discussion of Epiphenomenalism, the principle area of interest is that of human agency; the understanding of how our thoughts link with our movements. I shall assume that what Epiphenomenalism questions, is the truth of the possibility of our thoughts ever making a difference to our movement. Normally, it would be reasonable to distinguish bodily movement that was a result of deliberation from all other causes, such as reflex for example; so for the purposes of this chapter, 'action' will refer to movement that is supposed to be caused by deliberation, and 'behaviour' to all other types. Hence, 'Agency' is regarded as being the distinguishing causal process of actions. If agency is an illusion as Epiphenomenalism suggests, then all movement is 'behaviour' in

the above sense: just as Mercury behaves in certain ways in certain situations, such as expanding when heated, without there being a commitment to Mercury *intending* to expand in those situations, so too would people be said to behave in certain ways in certain situations, such as seeking medical assistance when ill, without there being any implication that the mind initiated such behaviour.

There is an awkwardness in talking about 'Minds' and 'Bodies', without wishing to appear to prejudge the relation between the two. It is undeniable that such a natural division exists, but much hangs upon the basis of that division - whether they may be divided on the basis of ontological or epistemological considerations. In spite of the danger of seeming to treat the mind and the body as separate, whilst intending to say that they are in some important sense the same, I shall refer to the mind as the locus of all thinking, feeling and emoting, and the body as being the locus of action and behaviour in the above senses. Two ways of thinking about the mind - body relation will be addressed in the process of distinguishing between Strong and Weak Epiphenomenalism. For convenience, the questions will be posed thus;

firstly, "How does consciousness affect the brain?",

secondly, "How does consciousness arise within the brain?"

A liberty is taken by posing the questions in this manner. It is recognised that not all concepts that fall within the domain of the mental will be captured by a notion of 'consciousness'. Some mental states are commonly used in explanation of behaviour, even though the agent may not be conscious of behaving from that motive: one might think of 'moods' as having an influence on behaviour, as a case in point. However, it is primarily the idea that agency involves consciousness that excuses this omission, plus the hope that the general account of the relationship between consciousness and the brain, will be adequate for dealing with all aspects of the mind.

### Section One

At first glance, one wonders why there has been such discussion over a thesis that is as counter-intuitive as Epiphenomenalism. Can anything be more obvious than the efficacy of mental states on our behaviour? Surely it is without doubt, that the discomfort of having the sun shining in my eyes causes me to raise my hand to shade them. Nor could one happily deny that an act of running from a building was caused by a belief that it was on fire, coupled with strong desire not to be injured; and this desire made sense because we appreciate that injury is something that is worth avoiding. Were we to entertain the thesis seriously, then these are explanations that we should indeed doubt and deny. Clearly, whatever motive is required to treat Epiphenomenalism with gravity, it must be strong enough to overcome common-sense. It is the purpose of this first Section to provide that motivation.

There are a number of ways in which our intuitions may be unsettled. These may be divided usefully into three kinds. The first exploits what Flanagan observes is a habit of over-rating the

influence of consciousness<sup>1</sup>. The second purports to have some scientific basis for questioning the exact rôle of consciousness. The third, and perhaps the most compelling, appeals to our ability to imagine a creature who behaves as we do, but who lacks consciousness - a 'philosophic' Zombie. If it is possible to imagine a creature that looks like a person and responds like a person, yet lacks sensation, then it follows that our criteria for personhood do not include consciousness. If behaviour, in its broadest sense, is not affected by the presence or absence of consciousness, then the thesis of Epiphenomenalism is justified. These kinds will be dealt with in turn.

### Subsection 1 : Is consciousness overrated?

One reason for doubting the truth of Epiphenomenalism, is because we take it for granted that the mind is crucially involved in our actions. There is a strong temptation to characterise oneself in terms of one's thoughts and experiences. My personality is most strongly manifested when I am conscious, so it is hardly surprising that we take consciousness to be ubiquitous. This ubiquity is unwarranted, however. Much of our behaviour can be accomplished without our being consciously aware of it. Having accepted that our awareness is not necessary in all cases, our confidence in it being necessary simpliciter, shrinks alarmingly. This is not improved by the recognition that a sensation may be registered after an action has been executed, as when the hand is jerked away from the fire, even before a feeling of pain is detected. Similarly, some intentional actions, like changing gear in a car, alter from being the sort of action that we have to concentrate on, to being the sort of behaviour that we can perform whilst devoting our conscious attention to another activity, such as wondering which would be the quickest route from Cannonmills to the Grange on a Saturday afternoon.

'Learning' also seems to be an activity that we can perform unconsciously, as exploited by many products claiming to be able to impart knowledge whilst you are asleep. Certainly, it is a fallacy to infer from the fact that some acts do not require consciousness, to the conclusion that none of them do; but having upset the simplistic identification of consciousness with agency, we require some criteria with which to ascertain just how much influence consciousness has. In spite of centuries of philosophic effort, these criteria are still absent; their absence permits Epiphenomenalism to insinuate itself into our thoughts.

One might chastise those who panic in the face of the above suggestions. After all, these are only at the level of 'common-sense', and no-one in their right mind pays heed to those sorts of consideration. What we need are 'facts'; science ought to be able to handle this. Yet, if we look to science to decide how much influence the mind has, then we faced with disappointment; some scientific evidence is taken to establish the truth of Epiphenomenalism. One

---

<sup>1</sup> Flanagan. 1992. P 133.

might suppose that Evolutionary theory would sanction our belief in agency. After all, the traits that become selected, are precisely those which contribute to the survival of the species. Since consciousness is present in so many genera, one may safely assume that it is there because it is useful. In which case, to doubt its importance is to ignore the mechanism of Natural Selection. However, the theory of Natural Selection fails to provide conclusive support for this opinion. For one thing, although selection equals usefulness, it is not the case that all traits are selected. Flanagan offers two such instances: the colour of blood and the shape of the chin. Some traits are linked with others in such a way that when these latter become selected, the former come with them willy-nilly. Although we find our awareness of pain useful for avoiding dangerous situations, there are many successful organisms that avoid danger without having recourse to such experiences. Having the property of consciousness is not necessary to survival, therefore. Only a species chauvinist would hold that it is more desirable to possess conscious mechanisms to avoid danger, rather than unconscious ones.

### Subsection 2: Scientific support for Weak<sub>E</sub>.

Further doubt is cast upon 'agency' by the phenomenon of 'Blindsight'. 'Blindsight' is a condition applied to individuals who claim that they lack visual experience in part of their field of sight, but who are apparently able to obtain information from that region. Physiologically, they have damaged the striate cortex (area V1). Psychologically, they are blind to aspects of their environment. This was initially quite puzzling to researchers: although the striate cortex is an important part of visual processing, the retina passes information to other regions of the brain via connections that bypass the striate cortex. Therefore, some visual information should be getting through to other parts of the nervous system. By encouraging patients to guess at what might be present in their blindfield, it was found that some visual information was available to them. What was clear, was that the patient had no conscious access to their blindfield. One Researcher notes:

Despite the demonstration of visual discriminative capacity, the subject characteristically will say that he does not 'see'. He is playing the experimenter's game.<sup>2</sup>

The behaviour of blindsighted people is such as to support the belief that they have some access to their blindfield. For instance, patients have been involved in an experiment using ambiguous words – such as 'Light', 'Bank' and 'Train'. Their task is to interpret these words. In their blindfield is displayed a word which favours one interpretation over another. In a significant number of cases, the patient's response matched the interpretation prompted by the displayed word. So, for example, where the word 'Light' is to be

---

<sup>2</sup> Weiskrantz. 1988. P 187.

interpreted, and the word 'Feather' is displayed (as opposed to 'Torch'), the interpretation would favour the connotation of weight rather than brightness. Further instances include the patients forming their hands appropriately when they reach out to touch objects that are in their blindfield. They are also able to say whether they are being shown an 'X' or an 'O' in their blindfield, at a level of correctness that is higher than would be expected were they to guess at random. The results from these findings, supports the idea that we can derive information from our environment without sensations. Such support puts pressure on those who have faith in the invaluable nature of consciousness.

In this regard, one may cite the findings of Benjamin Libet. In his experiment, Libet is concerned with the connection between activity in the area of the brain thought to serve hand movement, and the activity of the hand muscles. Subjects were asked to flex their hand spontaneously, though as soon as they felt the urge to flex their muscles, they were required to record the time of that decision. The results showed that awareness of an intention to flex occurred about three hundred ms *after* the onset of cortical activity, and about two hundred ms before muscle activation. Some subjects reported an awareness that they were about to flex their muscles, which was distinct from their deciding to do so. In short, the findings suggest that cortical activity precedes the conscious decision to flex muscle. As Libet says:

If the brain can initiate a voluntary act before the appearance of conscious intention, that is, if the initiation of the specific performance of the act is by unconscious processes, is there any role for conscious function?<sup>3</sup>

Libet's experiments have met with a mixed reaction, and there is some doubt about whether they show consciousness is always redundant. They are also taken as providing evidence that an unconscious initiation to flex can be stalled by a conscious decision to refrain from flexing. However, until we know more about the initiation of conscious thoughts, there is the suspicion that such a stalling decision itself may be the product of a prior unconscious process. In "Is Human Information Processing Conscious?", Max Velmans makes a case for the redundancy of consciousness in a third-person understanding of human behaviour. Using results from studies like the above investigations of Libet and Weiskrantz, Velmans proposes that when we process the information from our senses, much of this processing is hidden to introspection. He notes that, though we may sometimes be aware that we are 'thinking', 'remembering' or 'sensing' and so forth, this awareness does not invariably accompany these activities. Armstrong's example of the lorry driver, is a case where intentional actions were performed without there being any conscious accompaniment; an act is not disqualified from being intentional simply because we had not consciously formed the intention to carry it out. It may also be added, that we are not conscious of the manner in which our mental

---

<sup>3</sup> Libet. Cited in Flanagan, 1992. P 136.

processes are performed; we are unaware of how stimuli are analysed and selected or what feats of computation are necessary before we are able to utter our thoughts. Velmans concludes:

...conscious contents that *follow* given forms of information processing cannot be thought of as *entering into* that processing...If consciousness does not enter into human information processing, then the very notion that some of this processing is "conscious" needs reexamination.<sup>4</sup>

It should be pointed out, that Velmans denies that he should be thought of as an Epiphenomenalist. He believes that consciousness is only epiphenomenal from the third person point of view - a viewpoint he associates with that of science. Importantly, the first and third person perspectives are taken as complimenting each other, being mutually irreducible. Hence though *he* does not see his arguments as entailing Epiphenomenalism tout court, it is clear that a Materialist should. His position is strongly contested<sup>5</sup>, but if nothing more, the above discussion should serve to make us more respectful toward those who consider Epiphenomenalism a possibility.

### Subsection 3: The Zombie possibility.

Finally, the apparent ease with which we can accept the possibility of Zombies, has been taken as providing conclusive evidence in favour of Epiphenomenalism. Our ability to imagine a Zombie is clearly assisted by the first two considerations, which shake our confidence in the omnipotence of consciousness. Intuitively, Zombies are creatures who behave just as we do, but who have no feelings accompanying that behaviour. Hence, when a Zombie writhes and says "Parting is such sweet sorrow.", although there would be no physical difference between it and a human who truly described their feelings thus, all the same, there would be the important difference between the two; that whereas the Zombie's utterance does not refer to any sensation being experienced, the human's utterance would refer to a conscious state of sadness.

To put the case in its most favourable light, consider as an instance, a real Zombie - such as my neighbour Mr Bond. If I believe that Mr Bond behaves no differently from any other person of my acquaintance, yet has none of the experiences that those people have, then I am led to concede that feelings have no rôle to play in an explanation of our actions. On one occasion, I saw Mr Bond throw a hammer to the ground angrily, swearing in a quite imaginative fashion, place his thumb in his mouth, begin moaning about the futility of do-it-yourself, and then make the observation that a competence in practical skills can in no way be construed as a measure of intelligence. If asked why Mr Bond was behaving in that manner, it would be natural to reply that it was because he had just hit his thumb with the hammer.

---

<sup>4</sup> Velmans. 1991. P 666.

<sup>5</sup> Amongst many others, see Block 1991, Dretske 1991 and Van Gulick 1991.

Seemingly, feelings need not add anything to our understanding of behaviour; Mr Bond is physically and behaviourally no different from ourselves. It follows that just as we do not need to refer to a Zombie's feelings in interpreting their actions, because they have no feelings, then so too, it is unnecessary to refer to our feelings in the interpretation of our actions. If we accept the Zombie thesis, then the physical nature of the explanation "He has hit his thumb" seems to be sufficient for an adequate understanding of his behaviour: and if we do accept the thesis, we must accept that the case for Epiphenomenalism has been made.

What makes Mr Bond so easy to live with, is the fact that we are all used to interpreting behaviour without knowing with which mental state it is associated – if any. Apart from my own case, I am not intimate with the sorts of sensations and thoughts that accompany other people's behaviour. Feelings and behaviour come apart naturally when we interpret the motives behind other people's actions. In many cases of day to day life, my response to other people need not require that I make any judgement about their feelings at that moment; such responses may be determined by a desire to be 'courteous', rather than 'sympathetic', for instance. When I act purely out of politeness, such as when I hold open the door to a house for a visitor, it is not because I am responding to a feeling that I believe they are experiencing; rather, it is upon the assumption that they want to pass through inside. So far as I am aware, there is no distinct feeling "I want to go through that doorway" by which the visitor could be motivated. Nor indeed, need the visitor be experiencing any feeling of desire for entering the house: it may be that they are completely absorbed by feelings of excitement at the prospect of their birthday the next day.

To this extent then, how I act toward others, will on occasion be determined by an interpretation of their behaviour, in which no reference to feelings need be made. Such an interpretation need not even take into consideration the conscious intentions of others; the fact that the visitor has not formed the conscious intention to walk inside, does not provide me with an excuse to close the door in their face. In these cases it seems quite natural to be able to separate the physical processes from the mental ones.

Having got the two apart, it remains to be seen whether there is any justification for putting them back together. There are a range of responses to the above observations, which palliate the injury they do to our sensibilities; some are more effective than others. The following section deals with two natural, but flawed, ripostes to the above.

## Section Two

Epiphenomenalism presents itself as a position that no-one wishes to accept. It is a pressing problem for Monist and Dualist alike; the presumption of its falsity is one of the few things upon which they are in accord. Arguments against Epiphenomenalism are

generally motivated by an aversion to scepticism in all its forms. Two points are worth considering in this respect. Firstly, if one is committed to the denial of agency, then most of our explanations concerning human behaviour will have to be abandoned. Mind-body interaction is too useful in making sense of human life. Without it, much of what we take for granted, much that we regard as irreducibly part of the world to be explained, would simply be dismissed as an illusion.<sup>6</sup> Secondly, 'explanation' as a thoughtful human activity, would simply not exist. The practice of hypothesising about the nature of the world, as well as the setting up of experiments to test those hypotheses, would lose its significance were it the case that the mind were epiphenomenal; ex hypothesi, our careful deliberations would be of no consequence to the actions we perform. Epiphenomenalism, if valid, makes the idea of 'accepting as true on the basis of cogitation' unintelligible. Unfortunately, these considerations do no more than emphasise what is at stake. As arguments they lack conviction, as the following responses demonstrate.

Subsection 1 : Defence of Agency on the basis of explanatory importance.

According to the first point, there is a contest between our beliefs about the activity of 'explanation', and our beliefs about mental causation; these beliefs cannot be made consistent with one another. Many philosophers argue against the possibility of Epiphenomenalism simply because 'explanation' is so embedded in our ways of thinking. In "Metaphysics and Mental Causation"<sup>7</sup>, Lynne Rudder Baker argues that the apparent cogency of Epiphenomenalism is due to a misguided adherence to a belief in Materialism (which she defines as the thesis that every property-instantiation supervenes on a physical property-instantiation), combined with a commitment to the principle of the causal closure of the physical (which she defines as the thesis that every physical property-instantiation that has a cause at 't' has a complete physical cause at 't'). Granted these theses, then it is easy to see how the physical properties of events elbow out any mental properties. Baker's response is to retain the significance of the mental at the expense of metaphysics:

I think that our grounds for the claims that reasons sometimes explain behaviour are much stronger than any grounds for a metaphysical premiss that would lead to a contrary conclusion.<sup>8</sup>

Her views are echoed by Tyler Burge:

The probity of mentalistic causal explanation is deeper than the metaphysical considerations that call it into question.<sup>9</sup>

---

<sup>6</sup> See Burge 1993. P 117.

<sup>7</sup> Baker. 1993.

<sup>8</sup> Baker. 1993. P 95.



What the above arguments offer, is a choice between abandoning our trust in an activity that has served us so well in dealing with our environment, or surrendering our faith in a conception of how our minds are instantiated in the world. It only takes a moment to compare the usefulness of the practice of 'explanation' with Materialism, to see that it will not be the latter that people turn to when faced with problems like: "What is the link between climatic change and carbon emissions?".

### Subsection 2 : Response - inconsistency in our veneration of patterns.

Even so, it is far from clear that the two theses can be so easily separated. As was suggested in the previous chapter, the processes of scientific discovery depend upon there being a link between the world and our senses that is more than simply a brute correlation. There is a danger of assuming that truth will emerge inevitably from our explanatory methods. The point I wish to make, is that we are not consistent in our application of this rule; there are some correlations that we would wish to deny were 'explanations', in spite of their being as well supported as correlations which we treat with respect - in particular, the one between mental and physical states. Ironically, this respect arguably stems from the metaphysical position we adopt upon the mind's place in the world.

Whilst it is true that the efficacy of mental states plays an explanatory rôle in our lives, such a claim only establishes why we should want to defend the idea of mental causation. There is a useful pattern between the way the mind works, and the way the world works. Our conscious states are reliable guides to our environment. We do not know as yet how 'consciousness' is linked to the world, but the gist of the above remarks is, we may as well assume that the link is such as to enable the mind to make a difference. However, not every pattern reflects a law of nature; some are arbitrary or purely coincidental. Nor is the case that everything that has the form of an explanation may be regarded as such. For instance, it is a feature of the rainfall pattern in the south-eastern corner of Britain, that the period of least precipitation is during the months of June and July. However, it would be absurd to assume that rainfall is affected by clouds having an aversion to forming during any month beginning with the letters 'Ju'.

Doubtless less artificial patterns do exist, but it is erroneous to argue from their existence, to their being an instance of some law of nature. For example, an old farmer I knew, held that the sighting of a crescent moon, whose back lay facing the earth, was a sign of fine weather to come. The reason why, was that the rain was held in the dish formed by the two points of the crescent. Whether or not he accepted the explanation, it is certain that he believed in the truth of the assertion: he laid his plans according to his expectation of

---

<sup>9</sup> Burge. 1993. Pp 117-118.

dry weather, on the basis of seeing the moon in that state. There is nothing in our understanding of life that rules out a priori the existence of coincidence between events. The problem with patterns, is that they are a correlation between events. It is not obvious whether the relation is simply coincidental, whether the relation is necessary, or if contingent, which way the causal dependency lies between the events; that is to say, whether the events are ordered merely consecutively, or whether one is a consequence of the other. The following examples illustrate the danger of treating a pattern as if it were ipso facto determined by underlying causal laws.

Some patterns between events emerge through chance. If I notice that every time I shake hands with a sweep, the following day all my plans come to fruition, though every time I neglect this courtesy, my day is filled with frustration, then certainly there is a pattern that I have picked out. However unless it is a coincidence, there is much about science that would have to be changed in order to accommodate the law: "sweep-handshaking affects the behaviour of matter in such a way as is determined by the wishes of the individual who shakes the sweep's hand."

Some patterns between events have an element of necessity. If I study Genealogy in a half-witted sort of way, I may think I have stumbled upon an amazing truth about my family when I discover a pattern in the vocations of my Catholic forebears, that not a single one of the females went in for the Priesthood. I might also wonder why no-one has witnessed a planetary conjunction between Hesperus and Phosphorus.

Finally, some patterns exemplify laws of nature, but it is not always clear which of the events is the cause, and which the effect. I may note the correlation between two events, such as the emergence of the little wooden lady with an umbrella from the weather house, being closely followed by showers. If I conclude that her presence and absence must have some causal connection with the weather, I still have to decide if the connection runs from the lady's appearance causing the rain or vice versa. Though it would be false to think that she brought on the rain, there is nothing in the initial observation that establishes that fact: it is only in virtue of further knowledge about barometric pressure and weather systems associated with rain, in addition to knowledge about how pressure affects the mechanism in the weather-house, that brings out the relationship between the lady and the showers.

Epiphenomenalism does not deny that there is a pattern between mental and physical states; what it denies, is that the mind has any influence upon physical states. Instead, mental states are entirely determined by physical states. With the weather-house, we assume that the figures are predictors of weather rather than the other way round. Yet, if we do not understand the mechanism of the weather-house, then we have no justification for making that assumption. There are other patterns which people are reluctant to accept as evincing a real relationship between the mind and the world. Some of these are investigated and described in the practice of Astrology. Many dismiss the idea that the relative positions of

the planets affect our character, whilst being content to accept that our thoughts affect our behaviour. Astrology has a semblance to a scientific activity, and for many has great explanatory value: the correlations it utilises between planetary position and psychological make-up have been documented for many centuries. Yet, when it comes to respectability, Astrology is generally regarded as something of a pariah. Without wishing to be drawn into a debate upon the truth of Astrology, it does provide a parallel with our current position on the relation between the mental and the physical. If we are prepared to accept the latter, on the basis of its utility, then why not the former? Alternatively, perhaps our suspicion of the former should motivate us into being more scrupulous about those beliefs we except as true, merely because it is convenient for us to do so.

Whilst we might think it astonishing that there is no connection between thought and movement, it is not in virtue of our having a sound understanding of agency, because we do not possess that understanding. There is no doubt that there is a discernible pattern between the workings of the mind and of the body, but in order to justify our assumption that there are causal links between the two, we require an understanding of the process by which the events are related. It is no objection to Epiphenomenalism that the truth of agency is underwritten by the rôle it plays in explaining human action, since by parallel considerations, we would be committed to accepting any theories that purported to explain related phenomena, such as 'Sweeps', 'Black cats', 'Astrology', 'Men in the Moon' and so forth.

### Subsection 3 : Is Epiphenomenalism an unprofitable form of scepticism?

A related objection to Epiphenomenalism, is one motivated by a disdain for scepticism in all its manifestations. Scepticism is sometimes regarded as form of despair, rather than a cogent position. It has the dubious distinction of being the only philosophical position for which no argument can be provided in support. This singular feature has lead some to think that it does not thereby constitute a threat; after all, how threatening is something that cannot be articulated? Yet, the problem here is not that the sceptic, qua person, is unable to justify their position, for in order to do so, they would have to ignore their own prescriptions in formulating that view. Rather, it is whether the content of their belief could be true, regardless of its being essentially impossible to state: and granted the difficulty in finding a satisfactory way of accounting for the relation between mind and body, there is no obvious reason why that belief could not be true.

Frustration with the problem has led some philosophers to denigrate the activity of attempting to find a solution. Burge writes:

...in my view, the worries about epiphenomenalism have an air of make-believe. It is much surer that epiphenomenalism is false than that the various assumptions...that have been thought to lead to it are true.<sup>10</sup>

Certainly, one can agree with Burge that the position is not very helpful, but insofar as scepticism is an affront to our powers of understanding, it is more satisfying to consider this possibility seriously; for it does at least serve to point up the incoherence in the way we think about these matters, which suggests a need to reconsider the options available. Any solution will contribute to the overall clarity of our comprehension of ourselves, and the world we inhabit. We will not benefit by dismissing Epiphenomenalism simply because it is perceived to be a philosophical cul-de-sac.

### Section Three

The idea of strong Epiphenomenalism gains support from two related sources: that we are able to imagine a creature physically identical to ourselves, but who lacks consciousness: that there is a difficulty in reconciling a number of separate everyday intuitions that we have about the nature of ourselves and the world. What relates these sources, is an ignorance of how physical matter could possibly instantiate mental states. If it is possible to undermine this support, the strong thesis will pose no particular threat to a Materialist. The following will attempt to discredit these views.

The first line of support for Strong<sub>E</sub> comes from the logical possibility of Zombies. Subsections 1 - 4 will discuss how much weight this support can bear (The treatment of 'conceivability' arguments in Chapter Six, will be relevant here.). To begin with, a recent argument will be presented attacking the idea of Zombies. Some points will be made in criticism of this attack. Instead, three considerations are offered suggesting that Zombies are more difficult to imagine than their supporters believe. The second line of support comes from a set of commonly-held intuitions. These will be presented in Subsection 5 and shown to be inconsistent. The remaining subsections argue for a re-evaluation of these intuitions, particularly in the light of the conclusion of Chapter Six.

#### Subsection 1 : Kirk and the logical possibility of Zombies.

Zombies, like most things, have their fans and detractors in the Philosophic community. David Chalmers is a fan, and the arguments in Dretske's "Naturalizing the Mind" draw the conclusion that Zombies are creatures with different evolutionary backgrounds to ourselves. Robert Kirk is a convert to the detractor camp, and is firmly of the opinion that Zombies are not even logically possible.

---

<sup>10</sup> Burge. 1993. P 102.

He provides compelling reasons for this conclusion<sup>11</sup>, but these seem to contain a flaw. The following is a brief presentation of his position.

If it is possible to be a creature physically identical to myself, but lacking consciousness, then consciousness must be a non-physical property - referred to as 'qualia'. Compare a Zombie, 'Z', with a conscious individual 'C'. Kirk suggests that 'friends of Zombies' regard 'C' as a compound of 'Z' and qualia. Hence, they are committed to the following belief: were 'Z' to possess qualia, then it would become conscious. However, this contradicts their definition of 'Z', as being a creature wholly unaffected by non-physical properties. Furthermore, Kirk notes that it is nonsense to say of such a compound, that it is sensitive to the subjective character of experiences; for the Zombie part of the compound is by definition insensitive, and it is unintelligible to talk of qualia as being *sensitive* to experiences - they are properties rather than performances. Since it is assumed that there is nothing more to being a conscious being than qualia tacked onto a Zombie shell, and since neither of these individually is capable of detecting differences between types of experience, 'friends of Zombies' are forced to conclude that *nothing* has such an ability. Yet, this is a contradiction of the obvious truth of our own ability to make such discriminations. Therefore, Kirk concludes, we are not compounds of 'Z' and qualia, and thus Zombies are not genuinely possible.

### Subsection 2 : Problems with Kirk's account.

The problem is, whether the 'friends of Zombies' should accept that we are such a compound. According to Kirk, they are committed to the following: if my Zombie twin were to acquire qualia:

...that should make him conscious; for according to them the only relevant difference between him and me is that he lacks qualia.<sup>12</sup>

This is to ignore a further difference between me and a Zombie. Zombies do not require qualia in order to thrive successfully. If we are deprived of one of our senses, then we are thereby handicapped: with any such loss, we do not then behave like a Zombie. It appears to be erroneous to think of Zombies as humans that lack consciousness; the behaviour of humans in such a situation bears no resemblance to how we fondly imagine Zombies conduct their lives - Zombies do not behave like unconscious humans. Zombies are not deficient in any sense that supports the idea of a compound. Zombies do not need consciousness; a Zombie with such a faculty would be a highly surprised individual, but certainly would not be a human in any sense that springs to mind.

Nor is it clear why the 'friends' are committed to such a view of humans. It is a questionable notion that a body becomes human

---

<sup>11</sup> Kirk. 1999. Pp 5-9.

<sup>12</sup> Kirk. 1999. P 5.

simply by attaching qualia to it. We are not passive with respect to our experiences; we do things with them. Qualia alone do not constitute the mind. Kirk understates the position of the 'friends' here, by suggesting that they would be prepared to accept the compound story, without insisting upon the addition of all the other mental properties that humans enjoy. Included amongst these properties are those that enable the practice of discrimination. The confounding of the logical possibility of Zombies depends upon the assertion, that elements of the compound are unable to detect differences in qualia. However, for the reason stated above, it is far from clear that this 'compound' accurately depicts the difference between Zombies and ourselves.

Part of the difficulty with this issue, is that there are two challenges to our intuitions. One concerns whether we could imagine that conscious states have no effect upon our lives. The other is how would we go about detecting Zombies. Clearly the first is harder to achieve than the second, since it is impossible to imagine my life being the same without conscious states. On the other hand, if asked how would one detect a Zombie, it is natural to concede that one would be at a loss as to how to go about it. After all, they are physically, and behaviourally alike us in all respects, so there is no physical criterion upon which we could base our judgements; which is say, that there are no criteria at all, since Zombies are exhaustively physical.

### Subsection 3 : How not to imagine a Zombie; as a counterfeit.

Just how easy is it to imagine a Zombie? It is worth noting that the question assumes that we are in full possession of the physical facts, and that possession of this knowledge still makes it possible that we can imagine such creatures. At the present, this is not so remarkable. Even without taking the argument in Chapter Six into consideration, were anyone to identify consciousness as being - for instance - some function of forty-hertz oscillations, it is not clear to us what the implications of this identification would be. We would not know enough about the terms involved in the equation to inspire confidence in our acceptance. We can imagine a toddler accepting the possibility that two objects might have the same amount of energy but different masses. So too it *seems* possible to us, that a creature could have those areas of activity in their brains currently associated with visual processing, for instance, without any accompanying sensations.

There are two aspects to the question that require scrutiny. What is the nature of this possibility of imagination, and can we be assured that what we imagine really fills the specification of a Zombie? Certainly, Zombies appear to be a logical possibility, since the expression 'conscious creature' does not mean 'physical creature'. In other words, Zombies do not fall within the same category as bachelors and unmarried males of a certain age. However, I would argue that this appearance is misleading. What

needs to be remembered here, is that the concept "Zombie" is parasitic upon the concept "Fully functioning conscious being". That is to say, hypothetically, that were there *no* creatures instantiating consciousness, then there would be no sense to be attached to the notion "Zombie"; just as were it the case that there was no concept of "Five pound notes", there would not be a concept of "Forged five pound notes". Zombies are described implicitly in terms of human properties. So, just as it would not make sense to employ the idea of a counterfeit 'x' without implicitly referring to the idea of a genuine 'x', so too, is it senseless to talk about Zombies without referring to the idea of human consciousness. They are a sort of counterfeit. Their conceivability is not so independent of consciousness as to make it clear that there is a strong logical distinction to be made.

#### Subsection 4 : Are Zombies nomologically possible?

This is not to deny the possibility of creatures that resembled humans physically, but who lacked sensation. What would be contentious, is whether this could be a nomological possibility. If we hold to the notion that there are physical properties that instantiate conscious states, then we would have to doubt that the creature really was composed from the same materials as ourselves. However, this 'doubt' has little force whilst we are uncertain about the nature of consciousness. It is similar to telling a child that they cannot imagine that the speed of light could be exceeded. Such an impossibility appears arbitrary when removed from the context of theories of light. An analogous arbitrariness hangs around the Zombie case; yet the discussion in the previous chapter that linked physical properties to sensations, suggests a nomological coherence between the physical and the mental. Of course, we are at liberty to question the premise, that the properties of objects persist through all possible worlds; but if the only sense of 'possibility' is that which is a manifestation of such scepticism, then the Zombie should disturb us as little – or as much – as the possibility that the earth is really made of bread crumbs.

#### Subsection 5 : The impossibility of imagining myself to be a Zombie.

There is also a danger that we are less than scrupulous in our imagination. The resemblance between a Zombie and a human is less than immediately suggests itself. It is salutary to note that when we think of instances of Zombies, they are typically other people. Our acquaintance with other people's feelings, is mediated by their behaviour. Given the ability for humans successfully to feign feelings, it is all too believable that for any given piece of behaviour, the feeling that we would normally associate with it, need not be the current feeling experienced by the person. Yet this is not to say that the person is experiencing nothing.

Yet, whilst it is easy at first sight for me to suspend judgement about whether a creature is a conscious human or a Zombie, what does not make sense, is to think that I could be a Zombie. We may distinguish between two senses of "Me"; between the sense of "Me" that persists though I am unconscious, as when asleep; and the sense which is contained in the idea of a personality, that is expressed through the complex of my dispositions, beliefs, and desires for instance. In this latter sense, it is impossible for me to think of myself as a Zombie. After all, what would I have to imagine in order to think of myself as a Zombie, not in the sense of the possibility that I should be resurrected and haunt the neighbourhood, but that the mental identity should be preserved by this imagining.

If I am unable to imagine myself as being a Zombie, what can we say of Mr Bond in this respect? Mr Bond cannot picture anything to himself; he does not have the faculty of imagination to utilise. Although I have no trouble describing his behaviour in the hammer example, there are other times when it is hard to believe both that he is a Zombie, and that what he says makes any sense at all. For instance, recently he said; "I'm really sorry to hear about the death of your dog; I realise that you were very close. Words can say only so much, and yes, it is hard to believe now, but the pain does disappear in time. Cry, you'll feel better for not holding it back. I know, I've been there." What, I ask myself, could he possibly know about 'sorrow', 'feeling close to something', 'incredulity', 'anguish' and 'relief'? Yet all these concepts are contained in his assertion that he knows what I am going through.

This appears to lead to a dilemma for believers in Zombies. My Zombie twin is physically identical to myself, behaves exactly as I do, but lacks consciousness. The following suggests that the three criteria above cannot be satisfied when dealing with an individual's relation to themselves. For, when providing an explanation for consciousness, either the physical is all there is, or there is something over and above the physical. According to the first lemma, if I am unable to imagine myself as being a Zombie, then anything physically identical with myself would be unable to imagine themselves as being a Zombie. And so, if there are Zombies, then there must be some significant physical difference between us, which is a contradiction of the initial premise of the Zombie argument.

According to the second lemma, although we are physically the same, I have some extra stuff that explains how it is that I cannot believe that I am a Zombie. Behaviourally, there would be a difference between myself and my Zombie twin; if it were because of this non-physical property, then the case for Epiphenomenalism fails, since there is something that makes a difference which is not accounted for in our physical make-up, and is having an effect upon my behaviour. So, if feelings are allowed to make a significant difference to our behaviour, then the validity of the idea of a creature physically identical with ourselves and indistinguishable from ourselves is negated: we should no more consider their



behaviour as being human than we should think a parrot is being conceited when it screams; "Who's a pretty boy then?" It seems that there is a far stronger link between consciousness and humanity to make the idea of a Zombie real enough to provide support for Strong<sub>E</sub>.

#### Subsection 6 : An alternative reason for accepting Strong<sub>E</sub>.

The other source of support for Strong<sub>E</sub> emerges in consideration of the following. One can think of at least four commonly held beliefs, which regarded separately seem straightforward enough. However, an incompatibility emerges as soon as an attempt is made to bring these beliefs together. The dilemma is whether to hold on to all beliefs and accept that agency is an illusion, or to abandon one or more of them, thereby going against some of our most entrenched views. These beliefs may be described as follows.

The first assumption for convenience will be referred to as the 'Principle of autonomy'. It is widely accepted that there is something special about our minds which distinguishes the mental from the physical; this is demonstrated by the apparent possibility of being able to imagine a mind independent from a body. No doubt this does not mean a great deal, since what may be imagined can in no way can be relied upon to vindicate metaphysical standpoints; but at least it is far from absurd to think of the two being separate. Indeed, this separation is implicit in our belief that in most cases, there is a sense of contingency concerning laws of nature, whereas psychological laws are more conceptual in character. As such, whereas it seems to make sense to say that it might have been the case that charged particles attracted particles of the same rather than the opposite charge, one could not accept the possibility that the rules of valid inference might have been different - such that affirming the consequent and denying the antecedent would be valid means of deduction. The facts concerning the behaviour of particles were, in part, a product of testing hypotheses. Our belief that deductive reasoning is trust-worthy, is based upon a complex mesh of beliefs. However, this complex does not include the following; that it has been observed of all cases of deduction to date, that all have yielded true conclusions from true premises by valid forms of inference. This is not to deny that we could use inductive reasoning, in some instances, to learn about ourselves. We may infer from our actions that we are in love, for example.

This difference in how the term 'cause' is employed, goes toward explaining why we accept the following suggestion. Certain salient features of the mind, such as the character of sensation, intentionality of thought, and the uniqueness of the self, are features that could not possibly be explained in terms of the sciences. For instance, were a complete physical description of the world possible, then there must be a description of someone's experiencing the scent of a rose. However, there appears to be no way that I could identify what the smell is like from this description

alone. Experiences, as well as other features of the mind, such as the thought of how the scent is similar to that of aniseed, seem to elude the scope of the scientific form of description.

The second assumption will be referred to as the 'Principle of interaction'. Although the belief in the discreteness of the mental and the physical is strongly held, it is equally beyond doubt that the two are connected in some fashion. The example of the student in the library given above, is an instance where we would ordinarily allow that there is interaction between the mental and the physical.

Coupled with these beliefs concerning the relation between the mental and the physical, there are a number of beliefs concerning the way the physical world works. One is that there are discernible patterns in the way events in the world are connected. This assumption will be referred to as the 'Principle of regularity'. These regularities are generally immutable, and as such, may be relied upon in the furthering of our goals. For instance, granted a thorough understanding of mechanics, I can load my barrow in such a way that the weight placed in front of the wheel, counter-balances that behind it, making the force required to lift the handles more or less according to how the weight is distributed with respect to the wheel; the nearer the front of the barrow, the easier it will be to lift. Were I ignorant of these laws, I would be less able to minimise the effort involved in my labours.

Further, that the sequences of events that make up these patterns, are not arbitrary. This assumption will be referred to as the 'Principle of consequence'. It assumes that there must be some property of the antecedent event that brings about, or makes more likely, the consequent event. It is by alluding to these properties that we are able to provide explanations of why one event is caused by another. For instance, if you introduce some diesel fuel into a robust chamber, increasing the pressure in that chamber leads to an explosion. This is because, diesel is a combination of a number of substances, which at ordinary pressure are bonded together. However, when the pressure is increased, these bonds weaken allowing the separation into its constituent parts. These react together with an explosive release of energy.

### Subsection 7 : The inconsistency of these intuitions.

The first assumption holds that the mental and the physical are ontologically separate; that, amongst other factors, there is an essential difference in the way the term 'Cause' is being used to refer to the links between mental events, from its use to refer to links between physical events. The second, that nonetheless, there are connections between the mental and the physical which, in some sense, are properly referred to as causal connections. The third and fourth assumptions concern causal connections within the physical domain, and by the first assumption, employ a distinct usage of the term 'cause'. To give an air of respectability to the conflict that lies between these intuitions, it is possible to set them out in the form of an argument. This - styled the 'Exclusion Argument' - was

originally used by R. Van Gulick as a response to criticisms of Anomalous Monism, though it may easily be adapted to exclude the physicalist vocabulary;

1. *Token Physicalism*: Every mental event-token (i.e. every event-token having mental properties) is identical with some physical event-token (i.e. some event-token having physical properties).
2. The causal powers of a physical event-token are completely determined by its physical properties.
3. *The Non-reducibility of the Mental*: Mental properties are neither identical with nor reducible to physical properties.
- ∴ 4. A mental event-token's mental properties do not even partially determine its causal powers (from 1,2,3).
- ∴ 5. Mental properties are not causally potent (from 4).
- ∴ 6. Mental properties are not causally relevant; they are epiphenomenal (from 5).<sup>13</sup>

Given the first two assumptions, it seems natural to ask "How does consciousness affect the brain?" On the one hand, we expect to understand the relationship between two events, according to the state of affairs of the antecedent event bringing about the consequent event, where the events occur within the same ontological domain. On the other hand, agency requires that it is possible to describe a relationship between events that occur in different ontological domains. What is it about the desire to study that caused the student's behaviour? That is, what are the properties of 'desire' that could affect the properties of the brain, in such a way as to produce the movement? It is possible to understand what links the desire with a belief; namely, that the desire would be best served by better illumination, and thereby with his rising and turning on the lights. It is becoming possible to explain how activity in the brain affects muscles, by which movements are performed. The sticking point is providing an explanation of where in two seemingly seamless processes, there is room for causal influence to pass from the mind to the brain, that is, where the properties of desire interact with the properties of the efferent neurones.

### Subsection 8 : The problem of analysing 'cause'.

If we attempt to articulate the nature of this causal link as though it were on a par with other phenomena in the world, then it seems we are committed to doing so according to one of the following models:

Firstly, that there are three separate types of cause; one that applies to the relation between mental events, one that applies to relations between physical events, and the third that applies to relations between the two. So events in the mind cause events in the intermediate domain, which in turn cause events in the brain. Whilst this model avoids direct links between the mind and brain, and

---

<sup>13</sup> Van Gulick. 1993c. P 236.

thereby retains the sense of causal integrity, it does so at the expense of having to explain how the links with the intermediate domain are made. Events within this domain, must be susceptible to both physical and mental causal influences, if they are to convey influence from the mind to the body. But their being indifferent to whether they are caused by mental or physical causes, renders the initial motivation for distinguishing the mental from the physical on the basis of causal autonomy completely groundless.

Secondly, that there is one notion of cause which relates properties both within the mental and the physical, and between the two, but each mental state has properties that are physically efficacious, and each physical state has properties that are mentally efficacious. This model faces the problem that there is no difference between saying this, and saying that there is ontologically one sort of thing with two types of property. Nor would it ever be clear whether a mental state is brought about in virtue of the mental properties of a mental event or by the mental properties of a physical event. By sharing out the causal character thus, any distinction between the mental and the physical is completely lost.

Thirdly, that mental causes are compatible with physical ones such that within a single string of events, some will be mental and some will be physical. For example, in a series of events  $E_1$  to  $E_{10}$  it might be that  $E_1$  to  $E_4$  are physical,  $E_5$  to  $E_7$  are mental, and  $E_8$  to  $E_{10}$  are physical. This model is problematic for the following reasons; either these types of event are discrete, in which case there is no connection between  $E_4$  and  $E_5$  that explains the occurrence of  $E_5$ ; or there is a causal relation between  $E_4$  and  $E_5$  and  $E_7$  and  $E_8$ , in which case we return to the problem of having to say which property allows this connection to take place? The idea that gaps exist in neuro-physiological processes is well supported, but it is improbable that these are where the mental can intervene. Either the gap must appear at arbitrary intervals, though this conflicts with our belief that agency exerts an influence when it chooses, rather than when it has the opportunity. Or the gap is affected by agency, that is, it stops the physical processing for a while; but this account merely raises once more the problems of how the mental can interfere with a physical process.

What the above is meant to show, is that unless we rethink our positions on our beliefs about the mind, the world, and causation, we can only protect our intuitions concerning the 'Principal of autonomy' if we are prepared to cast it adrift from being within the scope of our natural understanding. If that is done, then it will always be questionable whether the relation between consciousness and the brain is such that supports agency, or whether it is merely a stipulation purely for the sake of preserving an explanatory practice or intuition. So long as there is this doubt, then strong Epiphenomenalism will be able to exploit it; and it seems that so long as we hold to the ontological autonomy of the mental, the doubt will remain.

### Subsection 9 : Is the Principle of autonomy at fault?

One immediate issue that has to be addressed, is that so far it has been assumed that the only way to defeat Epiphenomenalism is to abandon the 'Principle of autonomy'. This strategy is far from obvious, and certainly is not entailed by the above considerations. All they demonstrate is the incompatibility of a number of intuitions, not that one of them is conspicuously false. There is some concern that the term 'physical' is not plainly understood; some philosophers<sup>14</sup> doubt whether it is possible to find anything that corresponds to its usage. It may be, therefore, that presuppositions may be smuggled into the debate via use of this term. Insofar as 'cause' inherits much of the physicalist dogma, it would be unfair to pick on the autonomy of the mental as the culprit without first having considered the possibility that it is the notion of 'cause' that is at fault.

Some have attempted to utilise Hume's observations about there being nothing more to causality - certainly no link-like entity - than constant conjunction, to side-step the problem of finding a link between the mental and the physical that is compatible with the principles of regularity and consequence. For instance, the relationship between my intention to act and my acting, is sufficient for a Humean to assert that a causal relation exists between the two. However unintelligible the link between the mind and the body, that fact alone does not sanction a denial of the claim that the two are causally linked, for there are no causal connections that can be known a priori.

It is not clear that this emendation succeeds. How can states of consciousness arise from a seemingly inert substance that is governed by contingent laws? How is it that through the physical process of Evolution, at some point animals begin to have consciousness when hitherto there was none? How is it that an embryo at some stage in its development becomes conscious? The task of providing answers to the nature of the events in question, appears to require more than a different notion of causation. As Madell points out:

No recourse to a Humean view of causality will lessen our feeling of the special and disturbing inexplicability of these events.<sup>15</sup>

### Subsection 10 : Conclusion.

The argument of Chapter Six was that the mysterious nature of the link between mind and body is a natural consequence of the way we sense and comprehend our environment. Granted the cogency of that argument, it is clear that the Principle of autonomy need not be upheld; the autonomy of the mind appears because our theories are in principle unable to bring together physical and mental properties. One might think that Strong<sub>E</sub> is only a problem for Dualists, for

---

<sup>14</sup> See Crane and Mellor 1990.

<sup>15</sup> Madell. 1988. P 141.

whom the Principle of autonomy is not negotiable. However, this would be to ignore the problems that face Materialists, as Crane wryly points out:

On the one hand, the original motivation for physicalism was the need to explain the place of mental causation in the physical world. On the other hand, physicalists have recently come to see the explanation of mental causation as one of their major problems...How can it be that physicalist theories still have a problem explaining something which their physicalism was intended to explain in the first place?<sup>16</sup>

The Materialist must confront the possibility that although there are physical properties that have a mental description, the result is that any sequence of events requires only the physical description in order to be understood. In this case, we have the problem that was posed to Davidson: for all our talk of mental properties of events, it is only in virtue of their physical properties that one event follows another. There is, therefore, no need to advert to the mind in our explanations. This form of Epiphenomenalism requires an answer.

#### Section Four

Although there is a causal relation between what we characterise as the mental and the physical, it requires some argument to demonstrate that the relation is bi-directional: that states of consciousness make a difference to our behaviour. Weak<sub>E</sub> denies this by asserting that consciousness might be either irrelevant, or redundant. It could be irrelevant insofar as it merely comments upon our lives without affecting them, much as a Newscaster reports upon the day's events without thereby having any influence upon what is happening in the world. Importantly, the Newscaster is only aware of what has happened, and so can be thought of as continuously lagging behind the action. That is, our behaviour has always just taken place before we become conscious of it, so any decision we make will be concerned with a state of affairs that has already been dealt with automatically. As a result, our decisions become irrelevant to our behaviour. This position is provided with support from the above comments upon the impotence of mental properties.

Conscious states can be redundant insofar as they are a waste product of the brain's activity. A bicycle pump becomes warm as a tyre is being inflated; the heat is not contributing to the pressure of the tyre, but is a consequence of the changes in pressure brought about by the action of pumping. Analogously, consciousness is like that heat; it contributes nothing to the life of the individual, because life would continue in the same fashion were it the case that, contingently, consciousness was not a consequence of neural activity.

---

<sup>16</sup> Crane. 1995. P 211.

These forms of Epiphenomenalism trade upon the idea that consciousness is surplus to the proper functioning of an organism, yet it is far from clear that the case can be made. Materialists are accused of being committed to the idea of causal closure of the physical, which in turn, denies all relevance of mental properties in explaining behaviour. Between them, Steward, Cummins and Achinstein demonstrate that there are serious problems with our formulations of scientific models of explanation. Their views are sketched in Subsections 2, 3 and 4 respectively. As a result, the fact that mental properties cannot figure in scientific explanations should not be taken to imply that they cannot figure in true explanations. Subsections 5 to 7 deal with the suggestion that the mind is redundant, by pointing out the importance of feelings and perceptions to our lives.

### Subsection 1 : Is consciousness irrelevant?

One consequence of something's being superfluous, is that any activity with which it is associated would be unaffected by its absence. That is to say, whether or not there were Newscasters, the world would pursue its course for whatever reasons one would care to posit. When thought about in this light, it appears quite false that the world would be indifferent to Newscasters, since their presence does make a difference in the way populations act and think. Certainly, life would go on without them, but this is not to say that life is unaffected by their presence or absence. (If the analogy was changed to 'News Analyst', then one would have an image that fits neatly into the 'Higher Order of Thought' theories, discussed in Chapter Four.) Similarly with consciousness: one has only to think of defects of consciousness, whether they be sensory, emotional or rational, to see that their loss would make a big difference to life.

This response leaves unanswered the problem raised at the end of Section Three: is it in virtue of their mental or physical properties that mental events have an influence? As C.D. Broad pointed out in the mid-twenties:

Epiphenomenalism asserts nothing positive about the qualities and relations of mental events, and it denies only one thing about them. It simply says that mental events either (a) do not function at all as cause-factors; or (b) that, if they do, they do so in virtue of their physiological characteristics and not in virtue of their mental characteristics.<sup>17</sup>

If it is possible to provide an explanation of the workings of the brain, in terms of its physical properties, the fact that some of these properties are responsible for conscious states seems irrelevant. (It should be noted that other ways of thinking about causation are available; in particular, there is much to be said in favour of the distinction Fred Dretske draws between 'triggering' and 'structuring' causes.<sup>18</sup>) I want to suggest that such a suspicion

---

<sup>17</sup> Broad. 1925. P 473.

<sup>18</sup> For further discussion upon this distinction, see pages 197-198.

arises from a too uncritical acceptance of available models of explanation, particularly the Deductive-Nomological (or D-N) model discussed in Chapter Six.

Subsection 2 : Steward's criticism of the 'Network' model of causation.

The purpose of Helen Steward's article is to show that one commonly held view of causation is incoherent. She acknowledges that it is a restatement of a criticism made by Davidson of Mill's analysis of 'cause'. Mill argued that when explaining an event, it is possible to refer to particular causes and background conditions. Consider for example, the pumping of air into a balloon causing it to inflate. The causes of the expansion can be regarded as a particular instance of the application of air to the balloon, combined with the properties of the balloon, such as its being elastic and air-tight. She characterises the Millian Network model thus:

...the causation of any particular effect is typically a matter of a number of simultaneously existing causally efficacious particulars, usually conceived of as a mix of events and so-called 'token' states.<sup>19</sup>

Steward denies that it is possible to make sense of this 'mix'. She presents two premises that, if true, entail the demise of the Network model. Firstly that there is no requirement for 'token states' to appear in an explanans; and secondly that 'events' and 'token states' belong to such disparate categories as to render unintelligible the idea that they could be mixed. These premises are defended in turn, as follows.

When we say that A causes B, we are not committed to which properties of A or B are causally relevant. This is not a feature of 'token states': whenever we indicate such a state - for instance 'elasticity' - there is no doubt that it refers to a causally relevant property. This casts doubt upon their eligibility to appear in singular causal claims. In addition, adverting to token states adds nothing to an explanation that is not already implicit in the mentioning of particular causes. 'Token states' would be better regarded as causally relevant conditions, rather than particular causes. She regards the attempt to construe them as a species of particular cause, as arising from a rigid adherence to the idea that to play a part in a causal sequence, requires an active contribution to the effect: it is not accepted that facts about conditions could be causally relevant.

Next she examines the idea that a 'whole cause' would be forthcoming from the kind of combination envisaged by the Network model. It is suggested that attempting such a formulation would be pragmatically impossible; the list of relevant conditions - those that have to obtain, as well as those that must not - would be endless. If such a combination is to constitute the 'whole cause' of an event,

---

<sup>19</sup> Steward. 1997. Pp 126-7.



then we will never be in a position to formulate the whole cause of anything.

The Network model operates in the background of our thinking about mental causation. If the validity of the model is questioned, then we should use it warily - in particular, when using it to support views upon the irrelevance of mental properties to explanations of our behaviour.

### Subsection 3 : Cummins and a critique of models of explanation.

Robert Cummins<sup>20</sup> observes that there are two strategies that may be employed in explanation; these are identified as 'Subsumption' and 'Analysis'. The former lends itself to those explanations dealing with changes. Theories involved with the understanding of such changes, he calls 'Transition theories'. If I need to explain why event<sub>1</sub> caused event<sub>2</sub>, the commonest way of accomplishing this, is to subsume the events beneath the appropriate theory. However, this leads to problems with some physical laws - he offers the example of the 'pendulum law'. Here the Subsumption strategy fails to discriminate between the period of the pendulum being caused by its length, and the length of the pendulum being caused by its period. There are also problems with explanations that involve dispositions, since Subsumption strategies only specify dispositions, without explaining why such properties exist.

The other strategy is concerned with the explanation of properties, rather than changes. It does this by analysing the nature of the object in question; breaking it down according to its component parts and structure. In the end, it appeals to a form of atomism, where elements of composition are taken as the building blocks of all material. Objects behave the way they do, because of their constituent parts: the properties of these elements are stipulated, in a sense, marking the limits of explanation. Cummins' particular interest is in forms of Psychological explanation, and the Analytical strategy is strikingly productive in this field. This should not blind one to the fact that such a strategy is equally successful when applied to other scientific disciplines. Because Subsumption prevails as a strategy in the physical sciences, and Analysis in Psychology, there is a tendency to consider the latter as being an explanation strategy *manqué*. This however is unwarranted, and adds to the prejudice against taking seriously any explanation that does not follow the Subsumption strategy. It also places pressure upon scientific disciplines to adopt such a programme, even though it was not congenial to the subject matter:

Forcing psychological explanation into the subsumptivist mold made it continuous with the rest of science only at a price of making it appear trivial or senseless.<sup>21</sup>

---

<sup>20</sup> Cummins. 1983.

<sup>21</sup> Cummins. 1983. P 27.

Not surprisingly therefore, since we are used to depending upon such strategies to provide an understanding of our world, we are disinclined to look to other models of explanation to further our comprehension. When investigating the nature of human motivation, we are predisposed to accept that the explanation provided by the Subsumption strategy is superior to any other - particularly those that refer to such recondite entities as mental states. Clearly it is possible for these strategies to work in tandem. What is galling, is the Subsumption strategy has been pushed into prominence by D-N models, to the detriment of Analysis. Cummins notes:

This consequence is not surprising: subsumptive explanation of the sort featured in transition theories is the only sort of explanation for which the model is initially plausible. Analytical explanation...can be given a deductive-nomological format, but the result is completely uninformative, obscuring rather than illuminating the nature of analytical explanation.<sup>22</sup>

#### Subsection 4 : Problems with trusting scientific models of explanation.

What makes this domination by D-N models so frustrating, is that their validity is far from obvious. Peter Achinstein<sup>23</sup> discusses various problems with the model, one of which is of particular interest to the present debate. Achinstein suggests that models of explanation must meet two requirements. The first is that no single sentence, or conjunction of sentences, in the explanans, should entail the explanandum; a query is not satisfied by a response that is derived from the query itself. For example, "Because the wood split", is not an adequate reply to the question, "Why did the wood split?" The other requirement is that it must be possible to determine on a priori grounds alone, whether the relation between the explanans and the explanandum is suitable. Accordingly, the only empirical element of an explanation, relates to the truth of the statements composing the explanans.

He goes on to apply these requirements to D-N models, and finds them wanting. It is possible to construct an explanation that conforms to the above model, where the explanans contains a law and a true singular statement, where these are deductively related to the explanandum, but where the resulting explanation is false. The following is just one example (He offers three others to demonstrate this point)<sup>24</sup>. It concerns the explanation of the death of an individual, Jones. A D-N account is offered, with true premises, followed by a true conclusion by a valid form of inference; however, the account fails to provide a true explanation of Jones' death. Jones swallows a large amount of arsenic. It is true that ingestion of arsenic causes death. Therefore, Jones died after having swallowed the poison. Actually, Jones died from a different cause, such as being hit by a car. Achinstein notes that these models may only be

---

<sup>22</sup> Cummins. 1983. Pp 7-8.

<sup>23</sup> Achinstein. 1983. Chapter 5.

<sup>24</sup> Achinstein. 1983. Pp 167-70.

rectified at the expense of violating the above requirements - that is, either by including in the explanans, a sentence that entails the explanandum, or by conceding that it is an empirical question whether or not the explanans statements correctly explain the explanandum. He concludes:

For this reason, I suggest, D-N models which attempt to provide sufficient conditions for correct explanations in such a way as to satisfy both of these requirements will not be successful.<sup>25</sup>

The D-N model is thus limited in its application; it can only offer potential explanations. It is only reliable to the extent that we possess a complete description of the events leading up to the explanandum. Otherwise, there is a real danger that the purported explanans will suggest a cause that does not apply. Nor is this fault restricted to D-N models; debate over whether or not there is a perfect model continues<sup>26</sup>. The moral is, that such models are unreliable guides where information on the explanandum is sketchy. There is a temptation to judge an explanation according to how it fits one of these models, but as we have seen, conformity does not guarantee truth. The debate over the relevance of mental properties is conducted against a background of ignorance, concerning how these properties might be instantiated. As argued in Chapter Six, they are of a nature that necessarily eludes our understanding. This ignorance, coupled with a blind faith in there being a satisfactory model of explanation, has made the irrelevance of the brain's mental properties more compelling than is justified. It may be reasonable to accept that Jones died of arsenic poisoning; likewise, it may be reasonable to suppose we have explanations of behaviour that leave mental properties out of the account. In actual fact, Jones died in a car crash. We may not conclude that there is a complete account of behaviour to be given in physical terms, merely on the basis that it conforms to our model of scientific explanation.

#### Subsection 5 : Is consciousness redundant?

But is consciousness redundant? It could be said that the idea of consciousness as 'waste product', is importantly different from that of 'Newscaster'. It is clear that an inability to remove waste, may result in harm to a system. What might account for the disability that a lack of consciousness occasions, may not be because of any intrinsic value of the conscious states themselves, but because these states are the body's way of getting rid of inhibiting substances. Muscles cannot function properly with an build up of lactic acid, so a system is provided to disperse such an accumulation. It might be suggested analogously, that the brain cannot function properly without discharging the waste formed by neurological processes. If the manner of discharging the waste took the form of conscious states of one sort or another, then the

---

<sup>25</sup> Achinstein. 1983. P 173.

<sup>26</sup> For instance, see Ruben 1990.

disabilities associated with the absence of such states, would be due to the build-up of inhibiting substances, caused by the mechanism breaking down, rather than being associated with an intrinsic value in having those conscious states.

This objection assumes that a waste product is essentially useless; but this is not necessarily true. So-called 'kitchen waste' is a valuable source of nutrition to plant life. Whilst it might be that states of consciousness were the by-product of neuronal activity, unless it can be demonstrated that these products are without any use at all, then the Epiphenomenalist cannot say we can get by without consciousness. Some biological systems have by-products which are utilised in other systems; the heat that is a by-product of muscle contraction, is used in a shivering mechanism to maintain body temperature. Why then, should consciousness be any different? The above considerations have shown that consciousness is useful, in the sense that without it we are at a disadvantage. So even were it granted for the sake of argument, that consciousness is a waste product, that concession alone does not justify regarding the mind as redundant.

#### Subsection 6 : The importance of 'feelings' to our choices of action.

We believe that our judgements influence our behaviour; that if faced with the choice between competing possibilities, the outcome is undetermined until our deliberations are completed; that the issue is determined by the feelings we have concerning the possibilities available to us. It is not that, regardless of what we think, the choice is determined entirely by the underlying physical processes. We address ourselves to the feelings that other people have. What makes a difference between two actions 'A' and 'B', may be their effect upon a person 'X'. For instance, when approaching an actor after a first night performance, the feelings that I want to engender determine the sorts of things I should say to them. Were I sympathetic to the actor, and I had to choose between the following two phrases; "Call yourself an actor? I've seen better performances from a wardrobe" or "Never before has Chekov's prose been so well served - you have done humanity an immeasurable service this evening", then it is to be hoped that I would chose the latter, because the first phrase is likely to make him feel upset, whereas the second should make him feel proud. Or suppose that I want to banish my feeling of sorrow, and I am acquainted with Oscar Wilde's observation that "One has to have a heart of stone to read the death of Little Nell without laughing", it is natural to say that I should read Dickens' "The Old Curiosity Shop". The explanation of the action, at the level where feelings are regarded as insignificant, cannot make any sense of the motives behind such choices.

Subsection 7 : The importance of 'perceptions' to our choices of action.

It is pertinent to note some remarks made by A.J. Marcel<sup>27</sup> concerning the behaviour of Blindsight patients. He suggests the following experiment. A glass of water is placed in the region of the blind field of a thirsty patient 'A'. From what we know about the past behaviour of Blindsight patients, we may assume that 'A' possesses sufficient information to be able to identify the glass. Marcel speculates how 'A' will behave, and offers three alternative courses of action. 'A' will either: act as if the glass is perfectly visible to him; or, reach out for the glass, but without knowing why he is performing that action; or, simply do nothing at all. It is Marcel's guess that the patient will do nothing. This belief is based upon the behaviour of patients who suffer from brain damage that results in their ignoring part of their body. Depending upon which side of the brain the damage occurs, it is found that the sufferer will only dress one half of themselves and neglect the other half. He goes on to suggest that we base decisions to interact with parts of our surroundings to the extent that we are phenomenally aware of those surroundings. This is simply because we do not normally form intentions to act unless there is a good reason - hence the reluctance of subjects to reach out for objects in their blindfield. If we lack awareness of part of our environment, 'x', then it is senseless to suppose we could intend to make 'x' the object of our action.

Subsection 8 : Is it a contingent truth that the mind is not Epiphenomenal?

To that extent, one can consider Weak<sub>E</sub> no threat to agency. However, in doing so, it would appear that the importance of consciousness is only a contingent matter. Hence, it is quite possible that there are creatures whose physical processes are capable of providing all the information that consciousness does for us, but for whom consciousness may serve no purpose. For them, it would be very much in the same category as our appendix is to us. The difference would be, that for these creatures, a defect in consciousness would not have the same harmful consequences as it does for us. Lest it be thought that one could hardly ignore conscious states, this does appear to be possible in some cases. It used to be thought that cats were colour-blind, for all the difference it made to their lives; however, it seems that cats possess colour vision, it is simply that they do not bother to utilise it. Somehow, such a refutation of Epiphenomenalism does not seem sufficiently robust.

However, this is to confuse the possibility of maintaining a biological system of some complexity in the absence of conscious states, with the possibility of maintaining *our* system in such a situation. It so happens that our information about the world is mediated initially by sensations. They constitute the raw data upon

---

<sup>27</sup> Marcel. 1988. P 146.

which intelligent processes operate to guide the organism away from danger and toward safety, and nourishment. Theoretically, such processes could operate upon entirely unconscious forms of data - this is, after all, what happens in some machines. Arguably, in order to process the data, it is necessary for there to be some sense in which the data is treated as being part of a unified system. When the data is in the form of sensory experiences, then these experiences themselves become part of the overall interest of the system. Because we are sapient animals with the intelligence to reflect upon the rôle feelings play in our lives, we are able to alter our behaviour according to how those feelings are valued.

Clearly, our behaviour is often only intelligible when an explanation includes reference to consciousness. It may be contingent that information about the world is carried in the form of conscious states rather than in inert media utilised by computers: but what is not contingent, is that being sentient, we could be indifferent to the presence or absence of consciousness.

### Section Five

If we deny the propriety of thinking of causation on the pull and push model, then is the Materialist left with any recognisable account of causation at all? To refer back to the two ways of thinking about the relation between the mind and the body<sup>28</sup>, if we abandon the first: "How does consciousness affect the brain?", then it is necessary to enquire whether the question "How does consciousness arise from the brain?" has a reasonable chance of being answered? This section is concerned with answers of this form, by which the Materialist programme can continue, without being accused of having abandoned any respectable notion of causation.

Such a way of posing this question, suggests that consciousness might be an emergent property of the brain. In recent literature, there are versions of Emergentism that claim to be compatible with a Materialist stance - for instance, presented by John Searle in "The Rediscovery of the Mind" - and some that are antithetical to Materialism - those of E.J.Lowe. Subsection 1 will outline a Materialist position on emergent properties, followed by a criticism of this in Subsection 2. Here an alternative is offered, though in Subsection 3, this too will be rejected. Finally, alternative ways of construing the concept of 'cause' will be offered in Subsection 4.

#### Subsection 1 : Searle and emergent properties.

In "Rediscovery of the Mind", Searle offers the following analysis of what it is for something to be an emergent property<sup>29</sup>. Consider for example, a system S, made up of elements a,b,c and d. There will be features of the system that are not, or are not

---

<sup>28</sup> Mentioned above on P 165.

<sup>29</sup> Searle. 1994. Pp 111-112.

necessarily, features of the individual elements. Thus, if the system were jug, then it would be a feature of the jug that it weighed five pounds, but it would not be a feature of any of the molecules of that jug that they weighed five pounds. He calls these 'system features'. There is a distinction between those system features which can be deduced from the elemental features - such as shape - and those which can be deduced from the causal interaction of those elements, which he designates 'causally emergent system features'. Examples of these are 'solidity', 'liquidity' and 'transparency'. He then goes on to distinguish two notions of emergence: emergent1 and emergent2. The former are realised in a number of natural objects, and it is his view that consciousness is an emergent1 property of grey-matter. The latter notion is defined as follows:

A feature *F* is emergent2 iff *F* is emergent1 and *F* has causal powers that cannot be explained by the causal interactions of a, b, c...<sup>30</sup>

Searle completely discounts the possibility of there being any instances of Emergent2 properties outside the thoughts of Philosophers. To make sense of these properties would be to contravene the transitivity of causality - the idea that a property generates more causal power than is contained within its elements.

### Subsection 2 : Lowe's criticism, and re-interpretation of emergent properties.

Lowe is unimpressed with this point, and urges that consciousness is an emergent2 property. One reason for having nothing to do with the monist line is because, in placing the autonomy of the mental at the epistemological level, there is no longer any sense in saying that the mind causes anything. This reduces facts about the mind to the level of social or economic facts, which he claims is deeply counter-intuitive. There is no solace in accepting that mental concepts have explanatory value, he points out, since that misses the importance of consciousness to us:

For what concerns me as an agent...is not how my action may be retrospectively *explained* but rather what, prospectively, will *cause* it - a particular conscious decision of mine, *qua* decision to move the arm just so, or some neural event in my brain quite inaccessible (at least under that description) to my consciousness.<sup>31</sup>

By means of analogies, Lowe illustrates how it might be thought that there are emergent2 properties in the world. Speculatively, one might imagine that within systems, there will be biases of one sort or another, which tend to influence the emergence of patterns of behaviour. To take an example of a system that exhibits such a pattern, without it being a case of intention,

---

<sup>30</sup> Searle. 1994. P 112.

<sup>31</sup> Lowe. 1993. P 633.

consider a tick's ability to find a host<sup>32</sup>. The tick is stimulated by photoreceptors, to climb to a vantage point. There, it is sensitive to the odour of Butyric acid which prompts the tick to drop toward this source. It then senses the temperature of the object upon which it has fallen, and if suitable, begins to extract fluids from that object.

Similarly, one might fashion a similar explanation to explain how biases in our grey matter, dispose us to act in certain ways, such that a character emerges from the sum of these biases. If you think of values as being those biases by a different name, then perhaps, it becomes easier to see how a 'personality' could emerge from our dispositions to behave in certain ways according to our innate sensibilities. This is certainly suggested by Lowe's analogy with the spider's web. The web constrains the spider's actions, forcing it to follow certain patterns of movement, but enabling it to do things that would otherwise be beyond the spider's powers, such as snare flies, and move across spaces to where the flies are. Just as the web is independent of the spider, and has powers which are separate from the spider's, so too, on this analogy, is the mind independent of the body and allows the body to perform many actions which would otherwise be unavailable to it.

Lowe cites the experiments of Libet in support of this view of how the mind interacts with the brain. These experiments, outlined above<sup>33</sup>, purport to show that out of a seemingly uncoordinated pattern of neural activity, there is a sudden coherence into concerted activity leading in action, which corresponds with the patient performing an act of will. Thus the act of will in some sense forms a template, like a web, which channels the activity of the neurones. Accordingly, the decision to move my hand does not initiate a sequence of neural events, but rather the decision serves to:

...coordinate a host of mutually independent neural events so as to induce them to converge upon one specific pattern of efferent activity rather than another, or none at all.<sup>34</sup>

### Subsection 3 : Difficulties with Lowe's position.

No doubt Searle would agree that these were examples of emergent properties. However, he might question whether they are strongly emergent; that is to say, that they have causal powers that are not deducible from the properties of the individual elements of grey matter. One difficulty facing Lowe, which he readily acknowledges, is that the example of the spider's web provides evidence of powers which are nevertheless physically explicable; that is, they are instances of emergent<sup>1</sup> properties. The problem is that the causal power of the mind is now explained in terms of some form of 'psychic' influence, and is no more how we would naïvely characterise agency, than the neuro-physical account which was

<sup>32</sup> This example has a fuller description in Kirk 1994b, Pp 106-108.

<sup>33</sup> See this chapter, Pp 168-169 above.

<sup>34</sup> Lowe. 1993. P 638.



chastised earlier. Given that the objection to Searle's account was that it failed to respect our intuitions, then granted that alluding to psychic powers does just the same, Lowe's account appears to be hoist with its own petard. Whatever turns out to be the truth of the matter, it is bound to be counter-intuitive, since we have been systematically baffled as to the nature of the mind, and our language and beliefs reflect this ignorance. For this reason, the nature of agency will not be obvious in any way that will allow it to be readily assimilated with what we think at present, so it should not put us off that any analysis appears uncouth.

More importantly, even were the possibility of psychic forces granted, so far, all that has been shown is how the mind influences the body; it remains to be explained how the body affects the mind, and it is not obvious that the analogies illustrate how such a relation could take place. It cannot be denied that consciousness is affected by activity in the brain; if there is damage to some areas of the cortex, then a person changes character. Whilst psychic forces might induce the convergence of neural activity, it is not clear how this connection between mind and brain illustrates how electrical forces affect the mind. Again, this suggestion of a gap allows the doubt about whether the psychic forces really do have the influence that is allotted to them, or whether their activity is strongly epiphenomenal.

#### Subsection 4 : Alternative ways of thinking about causes.

If there is a commitment to thinking of causes as always having some sort of 'oomph', then it is not surprising that 'consciousness' appears as an inadequate contender for that rôle. As Burge puts it:

Why should mental causes alter or interfere with the physical system if they do not materially consist in physical processes? Thinking that they must, surely depends upon thinking of mental causes on a physical model - as providing an extra 'bump' on the effect.<sup>35</sup>

The value of Lowe's examples is that they demonstrate situations where we would admit causation was taking place, but where it was influential in a more subtle way than that based upon the Deterministic 'billiard ball' model. It is not merely by determining the vectors of forces that one can explain action. One suggestive way of seeing how states of consciousness may be said to be just as influential, is by thinking of the reflection in the mirror of a person who is shaving. The mirror supplies vital information for the purposes of shaving safely. The movements of the shaver are not arbitrary, but are based upon those regions of the face which have yet to be shaved, and these areas are delineated by regarding the image in the mirror. Similarly, granted that states of awareness are as good a means as any of obtaining information about the world, then it is on the basis of the information contained therein,

---

<sup>35</sup> Burge. 1993. P 115.

that consciousness affects behaviour. Whereas a computer uses feedback loops - which we have no reason to suppose generate sensations - in order to function, animals employ mental states to gain information and form judgements. That there are many viable methods of sensing the same external properties, is demonstrated by patients who have devices that stimulate patterns on their skin analogous to patterns which would normally fall on their retinas<sup>36</sup>. These are so effective, that patients eventually describe the experience as 'seeing'.

In "Mental Events as Structuring Causes of Behaviour", Dretske distinguishes between 'triggering' and 'structuring' causes. To see how this distinction works, consider the operation of a lock that is opened by keying in a particular sequence of numbers. The Triggering cause of the lock's opening, is any individual keying operation; the Structuring cause is the configuration of levers and rollers that enable the device to function as a lock. These two types of cause, may be distinguished on the basis of their relation to their respective effects. Firstly, whereas Triggering causes are sufficient for their effects, Structuring causes are only necessary for the effect; the lock is structurally poised to open at the entry of the correct combination, but that may never take place. It is possible to think of these sorts of causes as constituting the background conditions for Triggering causes. Secondly, for each triggering cause there is one effect - one instance of 'keying' results in one 'opening'. However, once the structuring cause is in place, it is capable of producing many tokens of the same effect - the lock's mechanism is capable of effecting many openings and closings. One might add, that whilst it is possible to understand the triggering cause of an effect without thereby understanding the structuring cause, the converse does not hold: on understanding the structuring cause, one thereby knows the nature of the trigger required to bring the structure into play.

Dretske employs this distinction in the understanding of the relationship between biological and psychological explanations. He maintains that we should look to Biology to provide the triggering causes of our behaviour, and to Psychology for the structuring causes. It is important to his account that different states of the same object may combine as both kinds of cause to an effect. For instance, an object 'O' possesses properties  $P_1$  and  $P_2$ .  $P_1$  may provide the triggering cause of an effect 'E', where  $P_2$  provides the structuring cause of 'E'. Dretske believes that this helps to explain how the properties of the nervous system determine behaviour:

...there is reason to think that it is the brain's possession of certain electrical and chemical properties that is effective in triggering bodily behaviour while its relational (including historical) properties are relevantly involved in structuring the same output.<sup>37</sup>

<sup>36</sup> As mentioned above; Chapter Five, P 111.

<sup>37</sup> Dretske. 1993. P 128.

These two types of cause combine to provide illuminating explanations of animal behaviour. When a creature runs away from a predator, it is possible to examine both the triggering cause of this - the perception of the predator - and the structure which accounts for how the sighting and the response have become associated. Just as citing the triggering cause of the lock's opening is only of limited explanatory value, so too is it the case in explaining animal behaviour: much of our knowledge comes from thinking of 'cause' as doing more than having a 'bump'.<sup>38</sup>

#### Subsection 5 : Conclusion.

The plausibility of Epiphenomenalism stems in part from an ignorance of how the mind and body could be connected, and in part by a blind adherence to Deductive-nomological models of explanation. If we accept that the ignorance is to be expected, then we can turn our attention to other models of explanation - such as those suggested by Cummins and Dretske, for instance - and other ways of thinking about properties - such as those outlined by Searle and Lowe.

---

<sup>38</sup> See also Audi 1993 for a discussion of the distinction between 'dynamic' and 'sustaining' causation.

## Chapter Eight

### Section One

#### Subsection 1 : Summary.

The above thesis addressed the following question. How could one account for the distinctiveness of the mind, along with its interaction with the body? The three elements of the mind that resisted reduction, were observed to be essential for any intelligent system interacting with the world. However, because of the relation between processes of comprehension, and objects discerned by those processes, it is impossible to arrive at an understanding of our mental faculties on the same terms as our understanding of other features of the world. This consequence of the relation, explains why our acquaintance with our mind is so distinctive, and why the 'conceivability' arguments are so persuasive. Yet the impossibility of identification between mental and physical states, is taken to be epistemological rather than a metaphysical.

If mind and body are taken to be composed of the same substance, then the problem of interaction becomes easier to resolve. The belief in the possibility of the mind being epiphenomenal, was supported by the idea that science could provide a complete account of natural events, without having recourse to mental properties. However, it was pointed out that the models of explanation cannot always be guaranteed to give the true account; a false account is compatible with a valid form of explanation. This is especially important when not all the facts of a situation are known - as is the case concerning the physical nature of the mind.

#### Subsection 2 : Disclaimer.

It should be stressed, it is by no means the object of this thesis to disprove any other Philosophical position. The thesis was intended to provide grounds for being a Materialist. Agnostic Materialism is consistent with the facts as I am aware of them. I am reasonably confident that it is incompatible with other philosophies. That it might be correct, is another matter. Agnostic Materialism is only one way of accounting for the explanatory gap between mind and body. It may well be that our minds are wholly immaterial and are able to intervene with our bodies, just as the Dualist claims. Equally, there may be no such thing as a mind-independent reality. The limited ambition of this project, was to see whether the Materialist could come up with a reasonable suggestion as to why the intuitions we have are so paradoxical.

## Section Two

### Subsection 1 : Implications for topics in moral philosophy.

The remarks made above, on agency, personhood and the emotions, have an impact upon areas of Moral philosophy. In particular, the question of Free will remains to be addressed; also, the question of how weakness of will is possible; and how the comments on Emotion have a bearing upon moral and aesthetic theories.

Concerning Free will, such remarks are of particular relevance to the debate conducted between Frankfurt, Watson and Taylor<sup>1</sup>. Comments upon character made in Chapter Six, contribute to something like Frankfurt's position. Although our behaviour is determined by our dispositions to act, according to how we feel our values are best realised, that does not mean that we are victims of predetermined dispositional profile. The 'Higher-order' theories make sense of the idea that there is competition between our various dispositions. One should assume that in order to survive, there will be a practical limit to how much conflict there could be, so that we will display a fairly consistent set of values. However, it is possible that being aware of how we act and how we would like to be thought of, we can change our disposition profile. In this sense, the thought of competing desires gives some support to the idea that we have a choice. It is exercising this choice that is regarded as being the exercise of Free Will.

Nor need these values be purely subjective. Because our decisions are very often in response to other individuals or groups, it means that the circle of influence is broader than only our own sets of desires. It is possible that an expression of free will, manifests itself in decision making that integrates the individual's values so far as possible; both with the values they hold dear for themselves, but also according to the status they want to achieve with respect to other individuals and society at large.

The idea of a 'person' being more analogous to a team of dispositions rather than an individual soul, helps account for the phenomenon of weakness of will. It is familiar enough that one both has a desire for an object or action, with an equally strong aversion to it. Where the will is seen as unitary, as on many accounts of personal identity, this phenomenon is deeply puzzling. However, granted a character may be composed from various value systems, each competing for their own ends, then weakness of will becomes more easily understood. It is the competition between two desires, which come into conflict. Thus although one can see the need to perform some act, such as writing 'thank-you letters', there is a stronger impulse to pursue activities that do not further this need. My will does not appear to be strong enough to accomplish the task.

This raises questions about the operations of the will. Does it operate at all, or is it only an illusion - is Determinism true? The above account of personhood provides a straightforward

---

<sup>1</sup> This can be found in a collection of papers edited by Watson 1982.

explanation: there are two conflicting desires, one to satisfy an obligation, the other, perhaps, to avoid the embarrassment of writing forced and hypocritical letters. The more basic emotion of shame is competing with a more abstract desire to observe etiquette - and not surprisingly, the more fundamental the desire, the stronger it is. Similarly, this account of personhood has a lot to say about self-deception. It becomes possible to see how two character systems have no, or subverting, communication with each other, which would result in a genuine sense of the left hand not knowing what the right hand was doing.

The above remarks upon the emotions suggest that there is something corresponding to a natural property, to which we react when making judgements of morals and art. Although it is probably a highly disjunctive natural property, to which an individual's particular character set is sensitive, it nonetheless might account for why Emotivism is so strong a doctrine in these fields. This does not mean that a morality based upon such a system would be justified, but it would point to the rôle of the emotions vis à vis some objective property in the world. It might be possible to naturalise our moral and aesthetic judgements upon these grounds.

### Subsection 2 : Implications for topics in epistemology.

Because the proffered solution has risen partly in response to the 'knowledge' and 'conceivability' arguments, it is not surprising that it will have repercussions upon areas of epistemology. In particular, the analysis of human understanding has two interesting implications. The first concerns the status of a priori knowledge. It was suggested above that in order to acquire knowledge of ourselves and our surroundings, there has to be a process whereby sensory input is analysed, and output behaviour chosen. Any sort of process capable of bringing unity to such information, must operate according to certain rules: and these rules cannot be acquired by observing natural patterns, since the recognition of any pattern requires a prior ability to individuate objects or events.

The obvious consequence of this, is to regard some types of knowledge as prerequisite for cognitive acts: for instance, knowledge of how to employ identity and comparative relations, as well as the utilisation of rules governing entailment and transitivity. These skills imply the possession of a priori knowledge. In order to process sensory input, there is a fundamental requirement that the system be capable of the following tasks: be able to individuate objects; be able to compare two or more objects; determine whether they are identical in some respect, and if not, how they compare with each other. This sort of know-how is implicit in the above analysis of physical systems that provide an individual with information about the environment.

The second implication is related to the first, and concerns the existence of innate ideas. Just as a computer system needs to have a certain amount of hard-wired information before it can run any system, so too does human intelligence need to have some

information prior to its ability to process any input, and co-ordinate any output. As suggested in Chapter Six, any intelligent system must be supplied with processes that - for example - preserve unity, locate peripheral devices such as memory, and perform accessing functions. That this is a prerequisite for learning, is demonstrated by the existence of heuristic programs run on computers, that require a level of setting up before they can be executed. If the mind is to have a Materialist analysis, then the suggestion of Chapter Six, is that the notion of innate ideas will find application in an analysis of cognition: at present, this position is taken by Chomsky, in work concerning the acquisition of language. As such, the thesis appears generally to be sympathetic to a Kantian account, whilst being hostile to Empiricism.

### Bibliography

- Achinstein, P. 1983. *The Nature of Explanation*. New York: OUP.
- Alter, T. 1998. 'A Limited Defense of the Knowledge Argument'.  
*Philosophical Studies*, 90: 35-56.
- Armstrong, D.M. 1981. *The Nature of Mind and Other Essays*. Cornell University Press.
- Audi, R. 1993. 'Mental Causation: Sustaining and Dynamic.' In Heil and Mele, 1993: 53-74.
- Baker, L.R. 1993. 'Metaphysics and Mental Causation.' In Heil and Mele, 1993: 75-95.
- Block, N. 1998. 'Is Experiencing Just Representing?'  
*Philosophy and Phenomenological Research*, 58(3): 663-670.
1996. 'Mental Paint and Mental Latex'. In Villanueva 1996: 19-49.
1991. 'Evidence against epiphenomenalism.'  
*Behavioral and Brain Sciences*. 14.4: 670-672.
1990. 'Inverted Earth.' In Tomberlin 1990: 53-79.
- 1980a 'Are Absent Qualia Impossible?'  
*Philosophical Review*, 89: 257-274.
- 1980b. *Readings in Philosophy of Psychology*. Vol. One. (ed.).  
Cambridge MA: Harvard University Press.
1978. 'Troubles with Functionalism.' Reprinted in Lycan (ed.)  
1990: 444-468.
- Burge, T. 1995. 'Our Entitlement to Self-Knowledge.'  
*Aristotelian Society Proceedings*, 96: 91-116.
1993. 'Mind-Body Causation and Explanatory Practice.'  
In Heil and Mele 1993: 97-120.
1988. 'Individualism and Self-knowledge.'  
*Journal of Philosophy*, 85: 649-665.
1986. 'Individualism and Psychology.'  
*Philosophical Review*, 95: 3-46.
1979. 'Individualism and the Mental'.  
*Midwest Studies in Philosophy*, 4: 73-121.
- Brentano, F. 1874. *Psychology From an Empirical Standpoint*. Ed. Kraus.  
London:RKP..
- Broad, C.D. 1925. *The Mind and its Place in Nature*. London: RKP.
- Cartwright, N. 1996. 'Fundamentalism vs the Patchwork of Laws'.  
In Papineau 1996: 314-326.
- Cassam, Q. 1987. *Self-Knowledge*. (ed.). Oxford: OUP.
- Chalmers, D. 1996. *The Conscious Mind*. Oxford: OUP.
- Crane, T. 1995. 'Mental Causation.'  
*Aristotelian Society Supplement*, 69: 211-236.
- Crane, T. and Mellor, H. 1990. 'There is no question of physicalism.'  
*Mind*, 99: 185-206.
- Cummins, R. 1983. *The Nature of Psychological Explanation*. Cambridge MA: MIT
- Dancy, J. and Sosa, E. 1993. *Companion to Epistemology*. (eds.) Oxford: Basil Blackwell.
- Davidson, D. 1993. 'Thinking causes'. In Heil and Mele 1993: 3-17.



1987. 'Knowing one's own mind'. In Cassam 1987: 43-64.  
 1980. *Essays on Actions and Events*. Oxford: Clarendon Press.  
 1973. 'The Material Mind'. In Davidson 1980: 245-259.  
 1970. 'Mental Events'. In Davidson 1980: 207-227.
- Davies, M. and  
 Humphries, G.W.  
 Dennett, D. 1993. *Consciousness*. (eds.). Oxford: Basil Blackwell.  
 1991. *Consciousness Explained*. London: Allen Lane.  
 1987. *The Intentional stance*. Cambridge, MA: MIT.  
 1975. 'Why the Law of Effect Will Not Go Away.'  
*Journal of the Theory of Social Behavior*. 169-176.
- Descartes, R. 1954. *Discourse on the Method of rightly conducting one's  
 reason and seeking the truth in the sciences*. Translated by  
 Anscombe, E. and Geach, P. Nelson University Paperbacks.
- Dretske, F. 1995. *Naturalizing the Mind*. Cambridge MA: MIT.  
 1994. 'Misrepresentation'. In Stich and Warfield, 1994: 157-173.  
 1993a. 'Conscious Experience'. *Mind*, 102: 263-283.  
 1993b 'Mental Events as Structuring Causes of Behavior.'  
 In Heil and Mele 1993: 121-136.  
 1991. 'Conscious acts and their objects.'  
*Behavioral and Brain Sciences*. 14.4: 676-677.
- Egan, F. 1992. 'Individualism, Computation, and Perceptual Content'.  
*Mind*, 101: 443-459.  
 1991. 'Must Psychology be Individualistic?'  
*Philosophical Review*, 100: 179-203.
- Evnine, S. 1991. *Donald Davidson*. Oxford: Polity Press.  
 Flanagan, O. 1992. *Consciousness Reconsidered*. Cambridge, MA: MIT.  
 Fodor, J. 1991.a 'A Modal Argument for Narrow Content.'  
*Journal of Philosophy*, 88: 5-26.  
 1991.b 'Too hard for our kind of mind?'  
*London Review of Books*, 13(12): 12.  
 1990. *A Theory of Content*. Cambridge, MA: MIT.  
 1989. 'Making Mind Matter More'. *Philosophical Topics*, 17: 59-80.  
 1987. *Psychosemantics*. Cambridge, MA: MIT.  
 1975. *The Language of Thought*. New York: Thomas Y. Crowell.  
 1968. *Psychological Explanation*. New York: Random House.
- Frankfurt, H.G. 1982. 'Freedom of the will and the concept of a person.'  
 In Watson 1982: 81-95.
- Geach, P. 1957. *Mental Acts*. London: RKP.
- Grimm, R and  
 Merrill, D. 1988. *Contents of Thoughts*. (eds.). Tucson, Ariz:  
 University of Arizona press.
- Guttenplan, S. 1993. *The Companion to the Philosophy of Mind*. (ed.)  
 Oxford: Basil Blackwell.
- Hardin, C.L. 1988. *Color for Philosophers*. Indianapolis: Hackett.  
 Harman, G. 1990. 'The Intrinsic Quality of Experience'. In Tomberlin  
 1990: 31-52.
- Hart, W.D. 1988. *Engines of the Soul*. Cambridge: CUP.

- Heil, A. and Mele, F. 1993. *Mental Causation*. (eds.). Oxford: Clarendon Press.
- Hempel, C. 1962. 'Explanation in Science and in History'. In Ruben 1992: 17-41.
- Horgan, T. 1993. 'From Supervenience to Superdupervenience'.  
*Mind*, 102: 555-586.
- Jackson, F. 1993. 'Representative Realism'. In Dancy and Sosa 1993: 445-449.  
1986. 'What Mary Didn't Know.'  
*Journal of Philosophy*, 83: 291-295.  
1982. 'Epiphenomenal Qualia.'  
*Philosophical Quarterly*, 32: 127-136.
- Kant, I. 1909. *Fundamental Principles of the Metaphysics of Morals*.  
Trans T.K.Abbott.
- Kim, J. 1997. 'Does the Problem of Mental Causation Generalize?'.  
*Aristotelian Society Proceedings*, 97: 281-297.  
1993. 'Can Supervenience Save Anomalous Monism?'.  
In Heil and Mele 1993: 19-26.  
1992. 'Explanatory Realism, Causal Realism, and Explanatory  
Exclusion.' In Ruben 1992: 228-245.
- Kirk, R. 1999. 'Why There Could Not be Zombies.'  
*Aristotelian Society Supplement*, 99: 1-16.  
1994a. *Raw Feeling*. Oxford: Clarendon Press.  
1994b. 'The Trouble with Ultra-Externalism'.  
*Aristotelian Society Proceedings*, 94: 293-307.  
1991. 'Why Shouldn't We Be Able To Solve The Mind-Body  
Problem?' *Analysis*, 51: 17-23.
- Kripke, S. 1972. *Naming and Necessity*. Cambridge MA:  
Harvard University Press.
- Levine, J. 1998. 'Conceivability and the Metaphysics of Mind.'  
*Noûs*, 32: 449-480.  
1993. 'On Leaving Out What It's Like.' In Davies and  
Humphries 1993: 121-136.  
1983. 'Materialism and Qualia: The Explanatory Gap.'  
*Pacific Philosophical Quarterly*, 64: 354-361.
- Lewis, D.K. 1994. David Lewis; Reduction of Mind. In Guttenplan 1994: 412-430.  
1992. 'Causal Explanation.' In Ruben 1992: 182-206.  
1990. 'What Experience Teaches'. In Lycan 1990.  
1980. 'Mad Pain and Martian Pain'. In Block 1980a: .
- Loar, B. 1988. 'Social Content and Psychological Content'. In Grimm  
and Merrill 1988: 99-110.
- Lowe, E.J. 1996. *Subjects of Experience*. Cambridge: CUP.  
1993. 'The Causal Autonomy of the Mental.'  
*Mind*, 102: 629-644.  
1992. 'The Problem of Psychophysical Causation.'  
*Australasian Journal of Philosophy*, 70: 263-276.
- Lund, D.H. 1994. *Perception, Mind and Personal Identity*. Lanham, Maryland:  
University Press of America.
- Lycan, W. 1996. *Consciousness and Experience*. Cambridge MA: MIT.  
1990. *Mind and Cognition*. (ed.). Oxford: Basil Blackwell.

1987. *Consciousness*. Cambridge MA: MIT
- Mackie, J. 1986. *The Cement of the Universe*. Oxford: Clarendon Press.
- Marr, D. 1982. *Vision*. San Francisco: Freeman.
- McGinn, C. 1993a. 'Consciousness and Cosmology: Hyperdualism Ventilated.' In Davies and Humphries 1993: 155-177.
- 1993b. *Problems in Philosophy. The Limits of Inquiry*. Oxford: Basil Blackwell.
1991. *The Problem of Consciousness*. Oxford: Basil Blackwell.
- McLaughlin, B. 1989. 'Type Epiphenomenalism, Type Dualism, and the causal priority of the physical'. *Philosophical Perspectives*, 3: 109-135.
1992. 'Davidson's Response to Epiphenomenalism.' In Heil and Mele 1993: 27-40.
- Madell, G. 1997. 'Emotion and Feeling.' *Aristotelian Society Supplement*, 71: 147-162.
1988. *Mind and Materialism*. Edinburgh: Edinburgh University Press.
- Malcolm, N. 1982. 'The Conceivability of Mechanism'. In Watson 1982: 127-149
- Marcel, A.J. 1988. Phenomenal Experience and Functionalism. In Marcel and Bisiach 1988: 121-158.
- Marcel, A.J. and Bisiach, E. 1988. *Consciousness in Contemporary Science*. (eds.) New York: OUP.
- Mellor, H. 1995. *The Facts of Causation*. London: RKP.
- Nagel, T. 1986. *The View From Nowhere*. Oxford: OUP.
1979. *Mortal Questions*. Cambridge: CUP.
1974. 'What Is It Like to Be a Bat?' *Philosophic Review*, 83: 435-450.
- Nemirow, L. 1990. 'Physicalism and the Cognitive Rôle of Acquaintance.' In Lycan, 1990: 490-499
- Papineau, D. 1996. *The Philosophy of Science*. (ed) Oxford: OUP.
- Pettit, P. 1986. 'Broad-Minded Explanation and Psychology'. In Pettit and McDowell 1986: 17-58.
- Pettit, P. and McDowell, J. 1986. *Subject, Thought and Context*. Oxford: Clarendon Press.
- Putnam, H. 1975. *Mind, Language and Reality; Philosophical papers*. Vol. Two. Cambridge: CUP.
1974. 'The Meaning of 'Meaning''. In Putnam 1975: 215-271.
1973. 'Philosophy and our Mental Life'. In Putnam 1975: 291-303.
1967. 'The Nature of Mental States'. In Putnam 1975: 429-440.
- Rosenthal, D.M. 1994. 'Identity theories'. In Guttenplan 1994: 348-355.
1993. 'Thinking that One Thinks'. In Davies and Humphries, 1993: 197-223.
- Ruben, D-H. 1992. *Explanation* (ed.). Oxford: OUP.
1990. *Explaining Explanation*. London: RKP.
- Ryle, G. 1949. *The Concept of Mind*. London: Hutchinson.

- Schiffer, S. 1976. 'A Paradox of Desire.' *American Philosophical Quarterly*, 13: 195-203.
- Searle, J. 1994. *The Rediscovery of the Mind*. Cambridge, MA: MIT.
- Shapiro, L. 1993. 'Content, Kinds and Individualism in Marr's Theory of Vision'. *Philosophical Review*, 102: 489-513.
- Silvers, R. 1993. *Representations*. (ed.). Dordrecht: Kluwer.
- Skinner, B.F. 1953. *Science and Human Behavior*. New York: The Free Press.
- Smart, J.J.C. 1959. 'Sensations and brain processes.' *Philosophical Review*, 68: 141-156.
- Sober, E. 1985. 'Panglossian Functionalism and the Philosophy of Mind.' *Synthese*, 64: 165-193.
- Sosa, E. 1984. 'Mind-Body Interaction and Supervenient Causation', *Midwest Studies in Philosophy*, 9: 271-281.
1993. 'Davidson's Thinking Causes'. In Heil and Mele 1993: 41-50.
- Sprigge, T.L.S. 1983.. *The Vindication of Absolute Idealism*. Edinburgh: Edinburgh University Press.
1971. 'Final Causes.' *Aristotelian Society. Supplement*, 45: 125-140.
- Steward, H. 1997. 'On the Notion of Cause 'Philosophically Speaking''. *Aristotelian Society Proceedings*, 97: 125-140.
- Stich, S. 1992. 'What Is a Theory of Mental Representation?' *Mind*, 101: 243-261.
1982. *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge MA: MIT.
- Sturgeon, S. 1998. Physicalism and Overdetermination. *Mind*, 107: 411-432.
- Taylor, C. 1982. 'Responsibility for Self'. In Watson 1982: 111-126.
- Tomberlin, J.E. 1990. *Philosophical Perspectives, 4. Action Theory and Philosophy of Mind*. (ed.). Atascadero CA: Ridgeview.
- Turing, A. 1950. 'Computing Machinery and Intelligence.' *Mind*, 59: 433-460.
- Tye, M. 1996. *Ten Problems of Consciousness*. Cambridge MA: MIT.
1994. 'Naturalism and the Problem of Intentionality' *Midwest Studies In Philosophy*, 19: 122-142.
1992. 'Naturalism and the Mental'. *Mind*, 101: 421-441.
- Van Gulick, R. 1993.a 'Metaphysical Arguments For Internalism And Why They Don't Work'. In Silvers, 1993: 151-159.
- 1993.b 'Understanding the Phenomenal Mind: Are We All just Armadillos?' In Davies and Humphries 1993: 137-154.
- 1993.c Who's in Charge Here? And Who's doing All the Work? In Heil and Mele 1993: 233-256.
1991. 'Consciousness may still have a processing rôle to play.' *Behavioral and Brain Sciences*, 14.4: 699-700.
- Van Inwagen, P. 1982. The Incompatibility of Free will and Determinism. In Watson 1982: 46-58.
- Velmans, M. 1991. 'Is Human Information Processing Conscious?' *Behavioral and Brain Sciences*. 14.4: 651-669, 702-726.

- Walsh, D. 1998. 'Wide Content Individualism.'  
*Mind*, 107: 625-651.  
1996. 'Fitness and Function.'  
*The British Journal for the Philosophy of Science*, 47: 553-574.
- Watson, G. 1982. *Free Will*. (ed.). Oxford: OUP.  
1982. 'Free Agency'. In Watson 1982: 96-110.
- Weiskrantz, L. 1988. Neurophysiology of Vision and Memory.  
In Marcel and Bisiach 1988: 183-199.
- Wilson, J. 1999. 'How Superduper Does a Physicalist Supervenience Need  
to Be?' *The Philosophical Quarterly*, 49: 33-52.
- Wright, L. 1973. 'Functions'.  
*Philosophical Review*, 82: 139-168.